



Organisation
des Nations Unies
pour l'éducation,
la science et la culture



« Chaire UNESCO en prévention de
la radicalisation et de l'extrémisme violents »
Université de Sherbrooke, Université Concordia
et Université du Québec à Montréal

Ce que nous savons de l'évaluation de programmes de prévention de l'extrémisme violent

Une revue systématique méthodologique des évaluations
des programmes de prévention dans ce domaine

JUILLET 2022

PABLO MADRIAZA, DAVID MORIN, GHAYDA HASSAN, VIVEK VENKATESH,
MAUDE PLAUDE, CAROLINE DELI, MÉLINA GIRARD, LOÏC DUROCHER-CORFA,
RAPHAËL GRIJALVA-LAVALLÉE, KAREN POULIN



Organisation
des Nations Unies
pour l'éducation,
la science et la culture



« **Chaire UNESCO** en prévention de
la radicalisation et de l'extrémisme violents »
Université de Sherbrooke, Université Concordia
et Université du Québec à Montréal

POUR CITER LE RAPPORT

Madriaza, P., Morin, D., Hassan, G., Venkatesh, V., Plaudé, M., Deli, C., Girard, M., Durocher-Corfa, L., Grijalva-Lavallée, R., & Poulin, K. (2022). *Ce que nous savons de l'évaluation de programmes de prévention de l'extrémisme violent : Une revue systématique méthodologique des évaluations des programmes de prévention dans ce domaine*. Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (Chaire UNESCO-PREV).

Nous tenons à remercier sincèrement François Champagne, Iris Boyer et Michael King, qui ont non seulement lu attentivement ce rapport, mais nous ont également fait part de leurs précieuses recommandations afin que la qualité de ce rapport soit améliorée.

Financé par le
gouvernement du
Canada

Canada

LES AUTEURS



Pablo Madriaza

Professeur au Département de travail social de l'Université du Québec en Outaouais (UQO). Il a été coordonnateur scientifique de la Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (UNESCO-PREV) et coordonnateur général du Réseau des praticiens canadiens en prévention de la radicalisation et de l'extrémisme violents (RPC-PREV).



David Morin

Professeur titulaire à l'École de politique appliquée de l'Université de Sherbrooke et cotitulaire de la Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (UNESCO-PREV).



Ghayda Hassan

Psychologue clinicienne, professeure titulaire de psychologie clinique à l'Université du Québec à Montréal (UQAM) et cotitulaire de la Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (UNESCO-PREV).



Vivek Venkatesh

Vivek Venkatesh est cinéaste, musicien, conservateur et scientifique de l'apprentissage appliqué. Il mène des projets de recherche et de création à l'intersection de la pédagogie publique et de la littératie numérique critique. Il est cotitulaire de la Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents, directeur du Centre d'études sur l'apprentissage et la performance et professeur titulaire de pratiques inclusives en arts visuels à l'Université Concordia à Montréal.



Maude Plourde

Étudiante aux cycles supérieurs à l'École de criminologie de l'Université de Montréal et entreprenant des études spécifiquement liées aux enjeux actuels de sécurité intérieure.



Caroline Deli

Candidate au doctorat en criminologie à l'Université de Montréal et titulaire d'une maîtrise en neuropsychologie à l'Université d'Aix-Marseille, sa thèse porte sur les processus de radicalisation des Incels à travers l'étude des trajectoires de vie.



Mélina Girard

Étudiante à la maîtrise en criminologie à l'Université de Montréal et auxiliaire de recherche (UNESCO-PREV, RPC-PREV, CICC, DARC). Mélina est également co-fondatrice du Journal Universitaire de Criminologie/Academic Journal of Criminology (JUC/AJC) et du Regroupement Étudiant en Criminologie profil Analyse (RÉCA).



Loïc Durocher-Corfa

Doctorant en psychologie à l'Université du Québec à Montréal, et assistant de recherche pour la Chaire Unesco en prévention de la radicalisation et de l'extrémisme violent (UNESCO-PREV).



Raphaël Grijalva-Lavallée

Raphael est étudiant à la maîtrise en psychologie sociale sous la direction de la Pr. Roxane de la Sablonnière à l'Université de Montréal. Il s'intéresse aux phénomènes de la polarisation sociale et des conflits entre groupes ethniques en période de crise sociale. Plus spécifiquement, il s'intéresse à l'évolution des préjugés envers les personnes d'origine chinoise au Canada durant la pandémie de COVID-19.



Karen Poulin

Karen Poulin est diplômée d'une maîtrise en études politiques appliquées à l'Université de Sherbrooke. Elle s'intéresse aux enjeux féministes et à la condition féminine. Dans l'optique d'en apprendre toujours plus et d'élargir son éventail de connaissances, elle a été auxiliaire de recherche pour la Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violent (UNESCO-PREV).

Table des matières

Sommaire exécutif	7
Introduction	10
1. Lacunes théoriques et méthodologiques des études évaluatives de programmes faites sur le terrorisme et la radicalisation : état de la situation	13
2. Méthodologie pour l'élaboration de notre revue systématique	24
3. Résultats	28
3.1 Que savons-nous globalement des programmes évalués et de leurs études ?	28
3.1.1 Quelle est l'évolution des études par année ?	28
3.1.2 D'où viennent ces études ? (nombre d'études par pays et continent)	29
3.1.3 Littérature académique ou littérature grise ?	30
3.1.4 Quel type d'extrémisme les programmes évalués ciblent-ils ?	32
3.1.5 À quel niveau de prévention les programmes évalués interviennent-ils ?	33
3.1.6 Quelles sont les actions ciblées ?	36
3.1.7 Le financement a-t-il été rapporté ?	37
3.2 Qui sont les auteures et auteurs de ces études ?	38
3.2.1 Égalité des sexes	39
3.2.2 Origine géographique des auteures et auteurs : les États-Unis, le Royaume-Uni et l'Indonésie	39
3.2.3 Disciplines représentées dans le champ de l'évaluation	41
3.2.4 Omniprésence du milieu de la recherche dans les professions des auteures et auteurs	43
3.3 Comment ces programmes ont-ils été évalués ?	44
Une question de design méthodologique	44
3.3.1 Quel type d'évaluation ces études visaient-elles ?	44
3.3.2 Des évaluatrices et évaluateurs indépendants ?	47
3.3.3 Quelle approche globale a été utilisée ?	48
3.3.4 Combien d'évaluations ont utilisé des méthodes expérimentales ?	50
3.3.5 À quel moment les données ont-elles été recueillies ?	52
3.3.6 Savons-nous qui sont les participantes et participants qui ont pris part à ces études ?	54
3.3.7 Outils de collecte de données	57
3.3.8 Mesurer ou ne pas mesurer directement l'extrémisme violent : le type d'indicateur utilisé	58

3.4 Quels sont les limites, les difficultés et les conflits d'intérêts liés à l'évaluation des programmes de prévention de l'extrémisme violent ?	60
3.4.1 Études identifiant clairement les limites	61
3.4.2 Conflits d'intérêts réels et potentiels	61
3.4.3 Limites identifiées par les auteures et auteurs	62
3.5 Quelle est la qualité des méthodes utilisées ?	69
3.5.1 Quelle est la qualité des études qualitatives ?	69
3.5.2 Quelle est la qualité des études quantitatives descriptives ?	72
3.5.3 Quelle est la qualité des études expérimentales ?	75
3.5.4 Quelle est la qualité des études quasi expérimentales ?	78
3.5.5 Quelle est la qualité des études mixtes ?	81
3.5.6 La qualité des études s'est-elle améliorée au fil des ans ?	84
3.6 Deux études de cas : l'évaluation dans le contexte des programmes de prévention de l'extrême droite et évaluation des programmes de prévention en ligne	85
3.6.1 L'évaluation de programmes de prévention de l'extrême droite	86
3.6.2 Évaluation des programmes de prévention en ligne	88

Recommandations	91
Conclusion	94
Références	97
Annexe A : Liste des études retenues	112
Annexe B : Méthode complète	124

SOMMAIRE EXÉCUTIF

Ce rapport fait partie de la première étape du projet PREV-IMPACT Canada qui, soutenu par les Fonds pour la résilience communautaire du Centre canadien d'engagement communautaire et de prévention de la violence et Sécurité publique Canada, vise le développement et l'implantation de modèles canadiens d'évaluation des programmes en matière de prévention de l'extrémisme violent (PEV)³ et, ultimement, le renforcement des capacités des personnes intervenantes clés au Canada.

³ Dans le contexte de cette étude, nous utilisons les concepts de prévention de la radicalisation menant à la violence et prévention de l'extrémisme violent comme des synonymes, en privilégiant ce dernier et son acronyme (PEV).



Ce rapport vise à présenter le contexte, la méthodologie et les résultats d'une revue systématique portant sur les méthodologies des études évaluatives de programmes de PEV.

1.1 MÉTHODE

Dans cette revue systématique, nous avons inclus toutes les études basées sur des données probantes primaires qui ont évalué des programmes de PEV, publiées en anglais, français et espagnol jusqu'en décembre 2019. Pour ce faire, nous avons réalisé une recherche basée sur des mots clés dans 21 bases de données scientifiques et 228 sites Web d'organisations travaillant dans ce domaine. Dans une première étape, nous avons passé au crible les titres et les résumés des documents identifiés lors de la recherche documentaire initiale afin d'éliminer les études non admissibles. Chaque document a été révisé par deux personnes. Pour s'assurer que l'accord entre les personnes évaluateurs était adéquat pendant cette étape, un Kappa de Cohen a été calculé, avec un résultat très positif ($Kappa = 0,86$). Les études retenues pendant la première étape ont été révisées au complet lors d'une deuxième étape afin de confirmer qu'elles satisfaisaient tous les critères déterminés par cette recherche. Finalement, nous avons évalué la qualité méthodologique des documents scientifiques retenus par l'intermédiaire d'un outil spécialisé.

1.2 RÉSULTATS

Sur un total de 18 886 documents identifiés pendant l'étape initiale de cette recherche, nous avons retenu 219 études incluses dans 211 publications. La plupart de ces études ont évalué des programmes de PEV en Europe ($n = 98$), en Afrique ($n = 50$) et en Asie ($n = 42$). Seulement 18 études ont été identifiées en Amérique du Nord et 7, en Océanie. Malgré cette distribution continentale, les études se concentrent sur l'évaluation des programmes de quelques pays de la planète, dont le Royaume-Uni ($n = 50$), l'Indonésie ($n = 16$), les États-Unis ($n = 15$), les Pays-Bas ($n = 12$), le Kenya ($n = 9$) et le Pakistan ($n = 9$). Cette concentration se reflète également par les auteurs et auteures de ces études. La moitié d'entre eux sont originaires des États-Unis (24,2 %) et du Royaume-Uni (23,5 %), tandis que seulement 5 % sont des ressortissants de pays africains. Cette différence s'explique notamment par le fait que les auteurs et auteures occidentaux ont évalué des programmes dans des pays occidentaux et non occidentaux. L'exception est l'Indonésie, où la plupart des auteurs et auteures des études sont des ressortissants indonésiens.

L'an 2016 est un moment décisif quant au nombre d'études évaluatives publiées, à partir duquel s'amorce une tendance croissante. Si, pendant les années précédentes, la moyenne d'études publiées est de neuf par année, elle grimpe à 30 études à partir de 2016.

La majorité des programmes évalués ($n = 127$) ont ciblé l'extrémisme violent en général sans prendre en compte un type d'extrémisme particulier, suivis par ceux qui ciblent directement l'extrémisme violent en lien avec l'islamisme ($n = 84$) et, enfin, l'extrême droite ($n = 20$). Ces programmes évalués ont travaillé principalement sur un niveau de prévention primaire et primaire ciblé³ ($n = 136$), suivis par les programmes de prévention secondaire ($n = 61$) et tertiaire ($n = 46$). Ainsi, les résultats par type d'extrémisme et par niveau de prévention montrent que les programmes de PEV évalués semblent suivre une tendance au cours des dernières années vers un type de programme moins spécifique et plus universel.

³ Les programmes de prévention primaire ciblés, bien qu'universels, ciblent une communauté spécifique ou un secteur au complet de la population considérée à risque (par exemple, des jeunes ou des communautés musulmanes).

Deux types d'évaluation sont les plus souvent utilisés : l'évaluation d'impact (n=159) et l'évaluation du processus (n=110), ce qui montre l'intérêt principal de connaître davantage les effets de ces programmes que les facteurs qui expliquent leur mise en œuvre. De l'ensemble des études, 55,3 % d'entre elles ont eu recours à des méthodes quantitatives, dont la grande majorité étant des designs mixtes (41,6 %) et 43,8 % des études qui ont utilisé exclusivement des méthodes qualitatives. Cette revue systématique a également identifié 54 études quasi expérimentales et 6 études expérimentales. Ces dernières sont une nouveauté par rapport aux recensions précédentes. Cependant, elles semblent être plus adaptées pour l'évaluation de programmes plus universels (prévention primaire) que pour des programmes plus spécifiques (prévention tertiaire). Dans ce dernier cas, les obstacles méthodologiques et les questions éthiques sont des facteurs qui empêchent une telle évaluation.

Les recensions de la littérature dans ce domaine mentionnent quelques limites importantes. En premier, nous notons l'absence de designs à mesures répétées dans les évaluations de programmes, c'est-à-dire basés sur plusieurs observations des mêmes sujets, à au moins deux moments différents. Dans le cas de cette étude, seulement 22,4 % des études ont utilisé des mesures répétées. Même si ce pourcentage est limité, on observe une évolution croissante de ce type de design au cours des dernières années. Deuxièmement, l'utilisation de groupes témoins est aussi souvent signalée comme un défaut des études dans ce domaine. Nous avons identifié 20 études qui ont utilisé un groupe témoin. Troisièmement, une importante limite provient du fait qu'il est très difficile de mesurer directement l'extrémisme violent dans le cadre d'études évaluatives. En effet, nous notons que la majorité des études a utilisé exclusivement des indicateurs indirects⁴ (74 %), alors que très peu d'entre elles ont recouru exclusivement à des indicateurs directs (4,1 %) ; un cinquième des études a utilisé les deux.

L'un des enjeux de ce domaine est la qualité des méthodes utilisées pour l'évaluation de programmes de prévention de l'extrémisme violent. Nous avons utilisé le *Mixed Methods Appraisal Tool* (MMAT) comme outil d'évaluation. Cet outil permet d'évaluer différents designs méthodologiques : les études expérimentales, quasi expérimentales, quantitatives descriptives, qualitatives et mixtes. Très peu d'études ont satisfait tous les critères mesurés par le MMAT et, en moyenne, les résultats de chacune des méthodes évaluées sont modérés, les études quantitatives descriptives étant celles qui ont eu l'évaluation la plus faible. Le problème le plus transversal des études évaluatives sur le plan de la qualité est celui des limites dans la transparence des méthodes, c'est-à-dire la quantité d'informations disponibles et le niveau de détail de ces informations dans les sections méthodologiques. Les études quasi expérimentales, par exemple, manquaient souvent d'une information détaillée concernant la représentativité des échantillons. Pour leur part, les études expérimentales ont été moins bien évaluées en ce qui concerne la randomisation des personnes participantes entre le groupe témoin et le groupe de traitement et dans la comparabilité des groupes au début. En revanche, les études qui ont utilisé des méthodes qualitatives ont été les mieux évaluées, malgré le fait que la relation entre l'interprétation des résultats n'a pas toujours été corroborée par des informations concrètes.

1.3 CONCLUSION

L'évaluation dans ce domaine est difficile, certes, mais possible. L'absence d'évaluations nombreuses et robustes s'explique davantage par le caractère récent du domaine et par l'urgence d'agir de ces quinze dernières années. L'évaluation de ces programmes n'est pas, en fait, à quelques exceptions près, très différente de l'évaluation d'autres programmes complexes de prévention de la violence. Ceci explique peut-être l'augmentation considérable de ce type d'étude au cours des dernières années. Cette augmentation est encourageante, mais il reste encore plusieurs défis à relever, dont la qualité des études réalisées et le nombre d'études dans certains domaines spécifiques. C'est le cas, par exemple, des évaluations de programmes en ligne qui sont très peu nombreuses, souvent non indépendantes, et dont les résultats ne sont pas toujours très probants. C'est le cas également des évaluations de programmes de prévention de l'extrême droite qui sont aussi très peu nombreuses.

⁴ Les indicateurs qui ne mesurent pas directement la radicalisation, la radicalisation violente ou les sympathies pour l'extrémisme violent.



Introduction

Face à la montée des groupes extrémistes dans différentes régions du monde et à la diversification des formes, cibles et auteurs d'attentats, force est de constater que la réponse sécuritaire traditionnelle s'avère insuffisante et même parfois contre-productive. Ceci est le cas parce que les facteurs traditionnels qui expliquaient l'essor de ce phénomène, surtout initialement appréhendé sous l'angle du terrorisme dit *djihadiste* dans le contexte particulier des conflits au Moyen-Orient, n'ont pas suffi pour comprendre les nouvelles vagues d'attentat extrémiste ou terroriste en Occident. Les facteurs locaux et les trajectoires variées des individus ont fait alors l'objet d'une préoccupation croissante et d'une place accrue dans la littérature scientifique. Cette évolution s'est accompagnée de l'essor de nouvelles approches conceptuelles et pratiques. C'est ainsi que l'on a commencé à accorder davantage d'attention à la notion d'extrémisme violent et à le conceptualiser à partir du processus de radicalisation menant à la violence. C'est également ainsi que l'on s'est intéressé à la notion plus large de prévention de l'extrémisme violent⁵. L'idée s'impose donc progressivement selon laquelle l'extrémisme violent peut également être prévenu par d'autres moyens que la lutte traditionnelle contre le terrorisme et les approches sécuritaires. Bien que ces dernières demeurent un pilier important des stratégies de lutte contre le terrorisme et l'extrémisme, de nouvelles perspectives, notamment psychosociales, émergent comme approches prometteuses tant en prévention qu'en intervention. La contribution spécifique de ces approches réside dans leur vision de la prévention comme n'étant pas l'empêchement de commettre un acte violent ou un attentat, mais plutôt comme un travail en amont qui vise à réduire ou à supprimer les conditions de risque qui peuvent rendre l'individu plus vulnérable à la radicalisation violente, à l'extrémisme ou au terrorisme.

Ce tournant préventif a permis d'élargir le champ d'action des personnes intervenantes qui commencent à incorporer des approches et des outils empruntés à d'autres disciplines, en multipliant en même temps l'offre de services et de programmes de prévention. En Europe, le Radicalisation Awareness Network (RAN) a, par exemple, identifié plus de 200 programmes prometteurs en prévention dans cette région, en 2019. Au Canada, le Réseau des Praticiens Canadiens pour la Prévention de la Radicalisation et de l'Extrémisme Violents (RPC-PREV) a, quant à lui, identifié 26 programmes de prévention secondaire et tertiaire au Canada, en 2020 (Hassan, Ousman, et al., 2020). Toutefois, cet essor n'a pas été appuyé ni accompagné par la construction d'une base conceptuelle et empirique clairement définie et rigoureusement délimitée, autant dans le cas des définitions des concepts utilisés que dans le cas des facteurs qui expliquent l'émergence du phénomène et, notamment, sur les solutions apportées. Heydemann soutient ainsi que « ce flou dans les frontières renforce les perceptions autour du

⁵ Dans le contexte de cette étude, nous utilisons les concepts de prévention de la radicalisation menant à la violence et prévention de l'extrémisme violent comme des synonymes, en privilégiant ce dernier et son acronyme (PEV).

CVE⁶ comme une catégorie fourre-tout » (2014, p. 3). Ce problème de définition qui touche à son tour la compréhension du phénomène de l'extrémisme violent elle-même : l'extrémisme violent étant alors définie aussi généralement que l'exprime Neumann, « ce qui se passe avant que la bombe n'explose » (2008, p. 4).

Dans ce contexte de flou conceptuel et empirique, l'analyse de la littérature mène à un constat récurrent : **il existe très peu de données probantes et de preuves tangibles concernant l'efficacité des mesures mises en place pour prévenir la radicalisation et l'extrémisme violents et très peu d'études ayant abordé la question ont été réalisées à partir d'un cadre méthodologique solide** (Baruch et al., 2018; Bellasio et al., 2018; Feddes et Gallucci, 2015; Gielen, 2017; Hassan, Brouillette-Alarie, Ousman, Kilinc, et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard, et al., 2021; Hirschi et Widmer, 2012; Romaniuk, 2015). Deux revues systématiques récentes qui portent sur des programmes évalués en prévention n'ont ainsi retenu que 48 études ayant dépassé un seuil minimal de qualité méthodologique sur une base de données initiale de plus de 15 000 documents (Hassan, Brouillette-Alarie, Ousman, Kilinc, et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard, et al., 2021).

Les chercheuses et chercheurs sont plutôt unanimes quant aux raisons qui expliquent ces lacunes. Il existe une difficulté conceptuelle, méthodologique et pratique pour mettre en place ces évaluations. Ce faisant, les recherches évaluatives existantes ont du mal à évaluer les résultats des programmes (*outcomes*) se focalisant davantage sur le nombre d'actions entamées (*outputs*) (Bellasio, et al., 2018; Feddes et Gallucci, 2015; Gielen, 2017; Lindekilde, 2012b; Mastroe et Szmania, 2016). Le cadre conceptuel diffus fait qu'il est difficile pour de nombreuses personnes chercheuses et praticiennes de déterminer ce qui doit être mesuré ou quels indicateurs d'efficacité doivent être utilisés. Par ailleurs, la méthode expérimentale, qui est le standard d'évaluation dans plusieurs domaines, est difficilement applicable dans le cas de la prévention de l'extrémisme violent. Les actrices et acteurs du PVE-CVE (*Preventing violent extremism – counter violent extremism*) ont ainsi peu d'éléments pour orienter leurs actions et se fonder sur des données probantes, des pratiques prometteuses ou qui ont fait preuve d'efficacité dans ce domaine. Malgré ce déficit, il existe un certain consensus parmi les personnes chercheuses, intervenantes et preneuses de décision quant au besoin de développer des modèles d'évaluation adaptés à ce type de programmes. Pour les intervenantes et intervenants, les évaluations permettent d'améliorer leurs pratiques. Pour les chercheuses et chercheurs, elles permettent de mieux comprendre les mécanismes et les processus qui expliquent le succès ou l'échec des interventions. Finalement, pour les preneuses et preneurs de décision, les évaluations permettent de guider la politique publique et de rendre plus efficaces les sources limitées de financement.

Afin de combler ces lacunes, la Chaire UNESCO-PREV a développé le projet PREV-IMPACT Canada. Soutenu par le Fonds pour la résilience communautaire du Centre canadien d'engagement communautaire et de prévention de la violence et Sécurité publique Canada, le projet PREV-IMPACT Canada vise le développement et l'implantation de modèles canadiens d'évaluation des programmes en matière de prévention primaire, secondaire et tertiaire de l'extrémisme violent et, ultimement, le renforcement des capacités des intervenantes et intervenants clés au Canada. Le premier volet de ce projet concerne fondamentalement la recherche et a pour objectifs de :

- Documenter et comparer les stratégies et outils d'évaluation en prévention de l'extrémisme violent (PEV) au Canada et ailleurs sur la base de données probantes et des pratiques existantes ;
- Développer des modèles d'évaluation distincts et novateurs (modèles logiques, stratégies, outils, indicateurs, méthodologie) adaptés aux programmes locaux de prévention primaire, secondaire et tertiaire afin d'orienter les politiques et programmes en PVE au Canada ;
- Tester les modèles d'évaluation sur trois programmes canadiens de PVE.

⁶ Counter violent extremism

Ce rapport est la première étape de ce projet plus large. Il vise à présenter le contexte, la méthodologie et les résultats d'une revue systématique portant sur les méthodologies des études évaluatives de programmes en prévention primaire, secondaire et tertiaire de la radicalisation et de l'extrémisme violents, publiées en anglais, français et espagnol jusqu'en 2019.

CE RAPPORT EST DIVISÉ EN QUATRE PARTIES :

01

La première partie (1) de ce rapport résume l'état de la situation sur les faiblesses, les enjeux et les défis de l'implantation des études évaluatives dans ce domaine. Elle se veut un résumé de l'état de l'art de l'évaluation dans ce domaine et sera la base sur laquelle cette revue comparera les avancées dans l'évaluation des programmes de prévention de l'extrémisme violent. Cette première partie est divisée en trois sections qui présentent : les lacunes dans le domaine (1.1), les difficultés rencontrées par les personnes évaluatrices dans l'application de ces évaluations et les raisons qui expliquent ces difficultés (1.2) ainsi qu'une brève revue des recensions passées dans le domaine (1.3).

02

La deuxième partie (2) de ce rapport décrit brièvement la méthodologie utilisée et les limites de cette revue systématique. La méthodologie est décrite plus en détail dans les annexes de ce rapport.

03

La troisième partie de ce rapport (3) présente les principaux résultats de notre revue systématique, en six sections portant sur : la description générale des études d'évaluation sélectionnées et des programmes évalués (3.1) ; les auteurs de ces études (3.2) ; les designs méthodologiques utilisés (3.3) ; les limites de ces études (3.4) ; leur qualité méthodologique (3.5) et deux études de cas spécifiques concernant l'évaluation des programmes de prévention de l'extrême droite et des programmes en ligne (3.6).

04

Finalement, la dernière section de ce rapport présente une discussion autour des résultats ainsi que des recommandations en matière d'évaluation de programmes de prévention de l'extrémisme violent.





Lacunes théoriques et méthodologiques

DES ÉTUDES ÉVALUATIVES DE PROGRAMMES FAITES SUR LE TERRORISME ET L'EXTRÉMISME VIOLENT : ÉTAT DE LA SITUATION

Dans les sous-sections qui suivent, nous répondons à trois questions clés dans le domaine :

- 1) Quelles sont les lacunes identifiées dans la littérature scientifique sur l'évaluation des programmes en prévention de l'extrémisme violent (PEV) ? ;
- 2) Pourquoi est-il si difficile d'évaluer ces programmes ? et Quelles raisons peuvent expliquer ces difficultés ?
- 3) Que disent les précédentes recensions de la littérature sur les programmes évalués et leurs méthodologies d'évaluation ?

Cette section se termine par une brève conclusion concernant l'état de l'évaluation dans le domaine.

1.1 QUELLES SONT LES LACUNES IDENTIFIÉES DANS LA LITTÉRATURE SCIENTIFIQUE SUR L'ÉVALUATION DES PROGRAMMES EN PEV ?

Les réflexions et les questionnements concernant les faiblesses théoriques et méthodologiques de la recherche sur le terrorisme et l'extrémisme violent constituent, depuis l'origine de ce champ d'études, une préoccupation récurrente. Au début des années 1980, Schmid et Jongman (1988, cités dans Silke, 2001) notaient déjà qu'il existait des problèmes importants en termes méthodologiques au sein des études sur le terrorisme, associés notamment à la collecte et à l'analyse de données : 92 % des études utilisaient, par exemple, comme source principale des journaux et des documents publics. Silke (2001), pour sa part, a analysé les articles publiés entre 1995 et 2000 dans les deux revues scientifiques les plus citées dans ce domaine (*Terrorism and Political Violence* et *Studies in Conflict and Terrorism*). Il a dressé un constat similaire : il existait une surreprésentation des recherches utilisant comme source principale des données secondaires ouvertes, notamment des journaux et des documents publics, des sources qualitatives anecdotiques (peu d'entretiens et non structurés) et très peu d'analyses quantitatives. Comme suite aux attentats du 11 septembre 2001 aux États-Unis, Silke a répété l'analyse des études publiées entre les années 2002 et 2004 et a vu très peu d'avancement dans le domaine (cité dans Neumann et Kleinmann, 2013). En 2013, Neumann et Kleinmann ont conduit l'étude la plus complète sur la question, concernant, cette fois-ci, les études sur l'extrémisme violent publiées entre 1980 et 2010. Bien qu'ils aient constaté une nette amélioration dans le domaine, plus d'un tiers des études recensées manquaient de rigueur soit méthodologique (le processus), soit empirique (le type de données) et un pourcentage important de ces études basaient leurs conclusions sur des données secondaires. Plus récemment, Shuurman (2018) a conduit une autre étude pour évaluer l'état de la recherche sur le terrorisme, entre les années 2007 et 2016, sur l'ensemble des articles publiés dans neuf revues spécialisées qui ont utilisé des données primaires. Il a également constaté une amélioration, notamment en ce qui concerne l'utilisation de données primaires. Il confirme toutefois que le domaine n'est toujours pas consolidé en raison, notamment, du fait que la plupart des auteures et auteurs sont des contributeurs ponctuels au domaine.

Dans le cas des évaluations de programmes en prévention, une série de lacunes méthodologiques additionnelles a été identifiée par les chercheuses et chercheurs. Celles-ci peuvent être résumées par le manque de

systématisation, de cohérence et d'harmonisation des méthodologies, le plus souvent *ad hoc* et qui ne répondent pas aux critères scientifiques minimaux requis (Davey et al., 2019 ; Feddes et Gallucci, 2015 ; Lindekilde, 2012b ; Marret et al., 2017). À ce jour, par exemple, aucune étude évaluative expérimentale n'a été identifiée et très peu d'études quasi expérimentales (Bellasio et al., 2018 ; Feddes et Gallucci, 2015 ; Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021 ; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021 ; Mastroe et Szmania, 2016). Le manque de données empiriques est, en fait, une caractéristique récurrente de ces évaluations qui se sont concentrées davantage sur la description des activités entamées (*output evaluation*) plutôt que sur l'efficacité des mesures mises en place (Baruch et al., 2018 ; Marret et al., 2017 ; Romaniuk, 2015). En outre, bien que de nombreux programmes aient été conçus pour obtenir des résultats à long terme, les évaluations se sont limitées au court terme n'utilisant des modèles longitudinaux que dans des cas exceptionnels (Marret et al., 2017 ; Romaniuk, 2015). Le recours à des groupes témoins est également rare dans ce domaine, de sorte qu'il est difficile de savoir avec certitude si les effets positifs ou négatifs des programmes sont dus à l'intervention ou à d'autres facteurs concurrents (Marret et al., 2017). Dans certains cas, cette difficulté est due à l'accès à l'information. À titre d'exemple, Davey et ses collègues (2019) mentionnent que, dans le contexte de programmes de prévention en ligne, les évaluations ont largement tendance à essayer de mesurer la portée de l'intervention et de l'engagement plutôt que les changements d'attitude ou de comportement en raison, notamment, de l'information restreinte sur les plateformes de médias sociaux. Les analyses cout-bénéfice des programmes sont également peu nombreuses (Marret et al., 2017). Seuls 19 % des échantillons analysés par Feddes et Gallucci (2015) comportaient, par exemple, des analyses des coûts économiques des programmes⁷. Un dernier problème observé par les chercheuses et chercheurs réside à la fois dans le manque de transparence concernant la méthodologie et les sources d'information et dans le manque d'indépendance des personnes évaluatrices (Horgan et Braddock, 2010 ; Mastroe et Szmania, 2016 ; Williams et Kleinman, 2014). Par exemple, dans le cas de quelques programmes en prévention de la récidive, le taux de succès a été basé sur des sources gouvernementales sans expliciter les moyens utilisés pour mesurer l'impact de ces programmes (Horgan et Braddock, 2010). Dans d'autres cas, notamment dans le contexte d'initiatives en ligne non gouvernementales, Davey et ses collègues (2019) soulignent que les programmes font rarement l'objet d'une évaluation indépendante. Ce constat n'est toutefois pas corroboré par d'autres recherches qui se sont focalisées sur l'ensemble des programmes en prévention (Bellasio et al., 2018).

⁷ Les auteurs ne mentionnent toutefois pas si cette analyse comprenait une analyse coûts-bénéfices en tant que telle.

1.2 QUELLES RAISONS PEUVENT EXPLIQUER CES DIFFICULTÉS À ÉVALUER LES PROGRAMMES EN PEV ?

Le manque de données empiriques concernant l'efficacité des mesures de prévention mises en place s'explique principalement par la difficulté à mener des évaluations dans un domaine récent et en constante évolution. L'équipe d'Holmer, Bauman et Aryaiejad considère que ces difficultés peuvent être regroupées en deux catégories : « les défis analytiques, tels que l'établissement de la causalité, le traitement des variations contextuelles et l'élaboration d'indicateurs valables ; et les défis pratiques, tels que la collecte de données pertinentes et fiables » (2018, p. 4). Le diagnostic de l'équipe de Marret (2017) est assez semblable et ajoute deux autres lacunes : le manque de méthodologie standardisée pour évaluer ces programmes et le manque de connaissances nécessaires de la part des personnes intervenantes pour concevoir et réaliser des évaluations adaptées à leurs programmes. L'analyse la plus globale et complète de ces difficultés a été réalisée par Bellasio et ses collègues (2018) qui ont analysé les difficultés conceptuelles, méthodologiques et pratiques identifiées par les auteures et auteurs des documents scientifiques qu'ils ont retenus dans leur revue systématique. Ils résument ces difficultés en cinq catégories :

- 1) la complexité inhérente au domaine (concepts et définitions, aspects de sécurité, etc.) ;**
- 2) les défis liés à la mesurabilité des phénomènes du monde réel (rareté des événements) ;**
- 3) les défis liés au design méthodologique (difficultés à établir un lien de causalité ou absence de théories du changement) ;**
- 4) les difficultés pratiques associées aux évaluations elles-mêmes (difficulté d'accès aux informations, contraintes financières, etc.) ;**
- 5) les inconvénients et avantages de l'utilisation de méthodes spécifiques (Encadré 1).**

Encadré 1. Difficultés identifiées en évaluation de programmes de l'extrémisme violent

1) Complexité inhérente au domaine

- Groupe cible
- Parties prenantes
- Aspects sécuritaires
- Interventions
- Concepts et définitions

2) Défis liés à la mesurabilité des phénomènes du monde réel

- Rareté des événements et absence de mesures des effets
- Manque de méthode pour mesurer les effets
- Évaluation des effets à long terme
- Prise en compte des normes et des attentes sociales
- Prise en compte de l'exposition à une intervention

3) Défis liés au design méthodologique

- Difficultés à revendiquer la causalité et à mener des expériences et des quasi-expériences
- Défis liés à l'adoption d'une approche d'étude longitudinale
- Absence d'une théorie du changement

4) Difficultés pratiques associées aux évaluations

- Contraintes en matière de ressources
- Difficulté d'accès aux informations sur les interventions et leurs effets
- Accès aux données
- Difficultés liées à la taille de l'échantillon

5) Inconvénients et avantages de l'utilisation de méthodes spécifiques

- Contraintes des recherches basées sur des modèles
- Contraintes des outils de recherche
- Importance de la triangulation et forces des méthodes qualitatives

Source : Bellasio et al., 2018

En nous inspirant de cette catégorisation et en ajoutant les constats d'autres chercheuses et chercheurs qui ont travaillé sur l'évaluation dans ce domaine, il importe de décrire plus en détail ces difficultés. Celles-ci nous semblent pouvoir être ainsi classifiées en sept catégories :

- a. Les difficultés conceptuelles et associées à la compréhension du phénomène**
- b. Les difficultés concernant la conception et la mise en place de programmes de prévention**
- c. Les difficultés concernant les actrices et acteurs impliqués**
- d. Les difficultés concernant le financement**
- e. Les difficultés concernant le nombre limité de cas et l'accès aux données**
- f. Les difficultés méthodologiques**
- g. Les difficultés concernant la politisation du phénomène**

a. Les difficultés conceptuelles et associées à la compréhension du phénomène

Dans l'introduction de ce rapport, nous avons noté que l'on se heurtait d'emblée à une difficulté au moment même de penser à l'évaluation d'un programme. Il n'existe pas de définition consensuelle concernant ce qu'est l'extrémisme violent, comment le délimiter par rapport à d'autres phénomènes et quelles sont les données empiriques qui peuvent expliquer son émergence (Lindekilde, 2012b ; Mastroe et Szmania, 2016 ; Ris et Ernstorfer, 2017). Bien qu'au cours des dernières années, il y ait eu une amélioration de la compréhension du phénomène, notamment par la publication de plusieurs études empiriques, revues systématiques et méta-analyses⁸, le domaine de l'extrémisme violent reste marqué par les faiblesses empiriques déjà mentionnées (Hirschi et Widmer, 2012 ; Lindekilde, 2012b ; Ris et Ernstorfer, 2017). Dans le cas de l'extrême droite, Hirschi et Widmer (2012) mentionnent, par exemple, que la nature du problème a été définie avec des degrés de précision différents, qu'il existe des explications concurrentes et que ce grand nombre d'explications différentes a été suggéré comme étant un obstacle à l'évaluation. D'autres chercheuses et chercheurs mentionnent également que la compréhension limitée des comportements individuels des extrémistes violents et des liens entre eux, ainsi que les multiples trajectoires qui mènent à l'extrémisme violent, empêchent une conception précise des programmes de prévention, souvent conçus sans base empirique solide (Baruch et al., 2018 ; Ris et Ernstorfer, 2017).

Bien que ce problème conceptuel puisse sembler très éloigné des réalités du terrain et de la pratique, il a des implications concrètes à la fois pour la conception des programmes (comment expliquer et induire des changements de la part des bénéficiaires de ces programmes), pour leur mise en place (critères d'inclusion et de référence des individus à risque) et pour l'élaboration d'indicateurs permettant de mesurer le succès d'une action. Par exemple, dans le cas de la prévention tertiaire, la récidive est souvent citée comme le seul critère factuel à considérer (El-Said, 2015). Cependant, la délimitation conceptuelle et empirique du concept de récidive est elle-même également mise en question, notamment en ce qui concerne les programmes dits de « déradicalisation » (Horgan et Braddock, 2010). Que signifie exactement la récidive ? S'agit-il d'un nouveau passage à l'acte, le fait de rejoindre un groupe extrémiste ou le retour d'idées radicales ? Quelle échelle temporelle doit être considérée ? Quels indicateurs factuels et quelles sources d'information devraient être utilisés pour mesurer ces changements de comportement ? De leur côté, Ris et Ernstorfer (2017) soutiennent que cette insistance sur une définition claire du concept et sur le caractère exceptionnel de ce type de programme néglige le fait que la plupart des approches visant à prévenir l'extrémisme violent sont également fondées sur l'expérience acquise dans d'autres domaines et peuvent donc être évaluées à l'aide de critères d'évaluation souvent utilisés ailleurs.

b. Les difficultés concernant la conception et la mise en place de programmes de prévention

La diversité et la particularité des programmes de prévention constituent une autre difficulté pour l'évaluation et l'harmonisation des méthodes d'évaluation (Hirschi et Widmer, 2012 ; Lindekilde, 2012b ; Marret et al., 2017 ; Mastroe et Szmania, 2016). On constate une grande hétérogénéité et un manque de compréhension commune quant aux objectifs et aux approches de ces programmes qui pointent dans des directions aussi multiples que différentes (Chowdhury Fink et al., 2013 ; Lindekilde, 2012b ; Marret et al., 2017). Les programmes de prévention sont, en effet, souvent conçus pour s'adapter au contexte local et aux particularités des individus ou des groupes qui constituent la population cible de l'intervention. Dès lors, ils ne peuvent être évalués sur une base générale, mais à partir d'une appréciation différenciée des effets spécifiques locaux. L'efficacité de ces mesures est, par conséquent, fortement dépendante du contexte de leur mise en œuvre. Selon Mastroe et Szmania (2016), cette situation donne à penser que les programmes élaborés dans un secteur géographique particulier pourraient être difficiles à transférer dans d'autres régions. Certains chercheurs et chercheuses

⁸ Voir, par exemple, Gill, Clemmow, Hetzel, Rottweiler, Salman et al., 2020 ; Vergani, Iqbal, Ilbahar, et Barton, 2020 ; Wolfowicz, Litmanovitz, Weisburd, et Hasisi, 2019.

estiment qu'il existe de surcroît très peu de preuves et d'apprentissages consolidés de ce qui fonctionne, dans quel contexte et pour quel type de public (Gielen, 2017 ; Ris et Ernstorfer, 2017). Au-delà de ces apprentissages, les évaluations sont rarement pensées au moment de la conception, voire de la mise en œuvre, du programme et apparaissent souvent tardivement comme un processus extérieur (pour ne pas dire un corps étranger) et déconnecté de l'initiative de prévention. Enfin, dans la plupart des programmes, l'absence de théories du changement constitue une difficulté additionnelle pour les évaluations (Bellasio et al., 2018 ; Chowdhury Fink et al., 2013 ; Williams et Kleinman, 2014). Les théories du changement (Encadré 2) permettent notamment d'expliquer comment les activités planifiées dans un programme de prévention sont censées produire les effets souhaités et, le cas échéant, d'identifier les mécanismes qui sous-tendent ces actions. Bien qu'il ne palie que partiellement à l'absence d'un plan d'évaluation intégré dès le début du programme de prévention, le recours aux théories du changement peut être particulièrement utile afin d'orienter les évaluations et de choisir les indicateurs les plus pertinents pour la démarche.

Encadré 2. Théories du changement

Connell et Kubisch définissent les théories du changement « comme une étude systématique et cumulative des liens entre les activités, les résultats et les contextes de l'initiative » (1998, p. 2). Les théories du changement ont été conçues pour évaluer et prendre en compte les impacts multiniveaux et multidimensionnels d'interventions complexes qui, à l'époque où cette approche a été conçue, avaient des difficultés à relier l'action aux résultats et où les approches évaluatives existantes étaient considérées comme inadéquates ou inappropriées (Sullivan et Stewart, 2006). Ce modèle d'évaluation fait partie des approches théoriques de l'évaluation et repose sur l'idée que l'évaluateur ou l'évaluatrice doit contribuer à faire émerger une théorie d'action implicite inhérente à une intervention afin de définir ce qui doit se passer si la théorie est correcte (Sullivan et Stewart, 2006). Il s'agit notamment d'identifier les indicateurs de changement à court, moyen et long terme qui nous permettront de déterminer les éléments nécessaires pour établir un jugement évaluatif. Cette approche est utile pour améliorer la planification des programmes, faciliter la prise de décisions concernant les méthodes d'évaluation et réduire les problèmes d'attribution causale qui sont souvent le fléau des évaluations de ce type d'intervention (Mackenzie et Blamey, 2005).

c. Les difficultés concernant les actrices et acteurs impliqués

Les programmes de prévention étant complexes, ils ont souvent recours à un large éventail de parties prenantes. Ces programmes ont des objectifs et des mandats différents et peuvent avoir aussi des besoins divergents concernant l'évaluation (Bellasio et al., 2018 ; Chowdhury Fink et al., 2013). La diversité des actrices et acteurs exige, en premier lieu, une évaluation complexe qui tient compte de ces différentes perspectives, ce qui implique que la portée, le temps et le coût de l'évaluation peuvent augmenter. Cette diversité exige également que ces différents points de vue soient mis en avant non seulement sur le plan de la collecte de données, mais aussi en ce qui concerne les besoins spécifiques du milieu de la pratique. La question de savoir pour qui l'évaluation est faite devient donc une question fondamentale. Certains bailleurs de fonds privilégient une évaluation d'impact des programmes selon une logique binaire (ils fonctionnent ou ils ne fonctionnent pas) et s'attendent à ce que l'accent soit mis sur la communication des résultats du projet (Ris et Ernstorfer, 2017). L'intérêt de certains gouvernements à montrer des résultats positifs n'est pas seulement une source potentielle de conflit d'intérêts, mais renvoie aussi à une question éthique en matière d'évaluation (Horgan et Braddock, 2010 ; Lindekilde, 2012b ; Mastroe et Szmania, 2016). Dans les cas des intervenantes et intervenants, l'évaluation peut être perçue autant comme une opportunité que comme une contrainte (Clement et al., 2021). Dans le premier cas, l'évaluation est pensée comme un moyen d'améliorer leurs pratiques quotidiennes et, dans le deuxième, est vue comme un audit qui surveille, voire remet en question leur travail. Finalement, concernant les personnes usagères du programme, elles sont souvent perçues comme une source d'information, tout simplement. Les approches participatives sont toutefois rares dans le domaine. Cependant, comme c'est une limitation qui mène à un enjeu d'évaluation, les usagères et usagers doivent être impliqués dès le début du processus pour réduire le risque de biaiser les évaluations. En raison du potentiel stigmatisant des actions de prévention identifié par des évaluations précédentes (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021), notamment en ce qui concerne la communauté musulmane, les personnes usagères peuvent avoir une réticence à participer à ce type d'initiative et à répondre de façon transparente aux questions d'évaluation. Cette situation est encore plus complexe dans le cas des interventions obligatoires, au sein du système correctionnel, par exemple, et rattachées aux procédures judiciaires (Mastroe et Szmania, 2016).

Pour de nombreux chercheurs et chercheuses, la question n'est pas seulement de savoir pour qui l'évaluation est faite, mais aussi qui l'a faite (Horgan et Braddock, 2010 ; Marret et al., 2017 ; Mastroe et Szmania, 2016). D'une part, l'expertise en matière d'évaluation est limitée et les formations complémentaires sont nécessaires afin d'avoir des évaluations qui prennent en considération les particularités du domaine. D'autre part, les questions éthiques soulevées ci-dessus montrent également qu'il est nécessaire d'instaurer une culture de la transparence autant de la part des gouvernements que parmi les chercheuses et chercheurs pour s'assurer de la fiabilité des données communiquées (Horgan et Braddock, 2010 ; Marret et al., 2017 ; Mastroe et Szmania, 2016).

d. Les difficultés concernant le financement

Le financement de l'évaluation est une source de préoccupation constante. En 2013, Chowdhury Fink et ses collègues constataient qu'il était difficile d'obtenir du financement pour les évaluations des programmes et que les fonds alloués étaient souvent modestes. Ceci peut s'expliquer par les coûts des évaluations. Dans une étude internationale réalisée parallèlement à cette revue systématique, et qui a interrogé 57 personnes chercheuses et praticiennes concernant les enjeux de l'évaluation, les intervenantes et intervenants ont mentionné que l'évaluation des programmes pouvait coûter plus cher que le programme lui-même (Madriaza et al., 2021). Bellasio et ses collègues (2018) mentionnent que le manque de financement a un effet négatif sur la conception, la mise en œuvre et la qualité des évaluations réalisées. Le financement a un impact notamment en ce qui concerne les contraintes de temps. Celles-ci empêchent la collecte de données au moment le plus approprié, manquant ainsi des occasions de collecter des données pertinentes, diminuent le temps d'analyse et entravent la conception et la réalisation d'évaluations solides.

e. Les difficultés concernant le nombre limité de cas et l'accès aux données

Une difficulté souvent mentionnée par les chercheuses et chercheurs réside dans l'accès aux données et à l'information pertinente. Malgré la préoccupation considérable que le terrorisme et l'extrémisme violent suscitent dans certains pays occidentaux, il existe une faible incidence de cas autant sur le plan du nombre d'événements, tels que les attaques terroristes, que sur le plan du nombre d'individus recrutés par des groupes extrémistes (Lindekilde, 2012b ; Pistone et al., 2019). Dans le cas d'une évaluation de trois programmes de prévention au sein du système de probation français, connu pour être un foyer de recrutement et d'extrémisme violent, les chercheuses et chercheurs ont dû renoncer à l'évaluation de l'un des programmes et au recours à un groupe témoin en raison, notamment, du manque d'effectifs. Ces

personnes chercheuses se sont contentées d'un nombre limité de cas pour les autres évaluations (Madriaza et al., 2018b). Schuurman et Bakker (2016) ont évalué un programme de prévention de la récidive aux Pays-Bas. Bien qu'il s'agissait initialement d'une évaluation d'impact, elle s'est transformée en évaluation de processus pour la même raison, le programme ne prenant en charge que cinq individus au moment de l'évaluation. Comme Baruch et ses collègues (2018) l'évoquent, mener des expériences scientifiques avec un faible nombre de cas est particulièrement difficile dans un domaine politiquement sensible. Dans les cas d'interventions en ligne, le problème d'accès est souvent dû aux limites des plateformes en ce qui concerne le type d'information recueillie (Davey et al., 2019).

L'accès à l'information est également complexe en raison de la nature même du problème. L'accès au groupe cible et aux documents pertinents pour la mise en œuvre d'une mesure peut être considéré comme une information sensible, voire confidentielle, par exemple, en raison de procédures pénales en cours, du travail des services de renseignement ou parce que cette information est jugée importante pour la sécurité nationale (Hirschi et Widmer, 2012 ; Ris et Ernstorf, 2017). Les services sociaux ne donnent pas d'accès à l'information des bénéficiaires non plus pour des raisons déontologiques. L'information devient donc inaccessible.

f. Les difficultés méthodologiques

Les difficultés méthodologiques pour la mise en place des évaluations sont en partie le résultat de l'addition des difficultés mentionnées précédemment. Cependant, ces évaluations rencontrent d'emblée un défi qui n'est pas différent des autres domaines où l'accent a été mis sur la prévention, à savoir démontrer qu'un comportement ou une action n'a pas eu lieu (Holmer et al., 2018 ; Lindekilde, 2012b ; Madriaza et Ponsot, 2015 ; Mastroe et Szmania, 2016 ; Ris et Ernstorf, 2017). Le faible nombre de cas à considérer en matière d'extrémisme violent constitue à cet égard une difficulté supplémentaire. Dans d'autres domaines, tels que celui de la prévention du crime, on rencontre un vaste nombre de crimes qui sont commis et qui permettent de modéliser mathématiquement l'impact d'une mesure. Ce n'est pas le cas pour la prévention de l'extrémisme violent.

Outre ce problème fondamental, les chercheuses et chercheurs s'accordent sur le fait qu'il n'existe pas d'indicateurs clairs, cohérents et harmonisés pour mesurer l'impact d'une mesure de prévention (Baruch et al., 2018 ; Davey et al., 2019 ; Feddes et Gallucci, 2015 ; Horgan et Braddock, 2010 ; Lindekilde, 2012b ; Mastroe et Szmania, 2016 ; Romaniuk, 2015). Le succès est difficile à définir et à observer, même dans le cas, comme nous l'avons évoqué ci-dessus, de la réhabilitation des individus qui ont commis des actes terroristes. Ceci est

probablement dû à la difficulté à détecter et à mesurer les attributs de l'extrémisme violent (Baruch et al., 2018). Bien que les cas soient peu nombreux, il n'existe pas un seul profil ou une seule trajectoire d'extrémisme violent, ce qui rend potentiellement illimité le nombre d'indicateurs à utiliser. Dans le cas de l'extrême droite, Hirschi et Widmer (2012) considèrent qu'il est difficile de séparer clairement les attitudes de droite des autres attitudes, notamment celles qui sont plus latentes, même si les caractéristiques plus manifestes (tatouages, symboles, costumes, etc.) sont relativement faciles à étudier. Les changements, quant aux idées radicales, sont ainsi difficiles à utiliser comme mesures de performance. Les personnes chercheuses ont donc davantage tendance à utiliser des indicateurs indirects (Marret et al., 2017) qui sont associés théoriquement et non empiriquement aux idées extrémistes et aux comportements violents. De même, la relation entre les idées et les comportements demeure difficile à établir et à prouver (Holmer et al., 2018). Il est souvent compliqué de démontrer, d'une part, une relation de cause à effet entre l'intervention et les changements et, d'autre part, que ces changements ne sont pas le résultat de facteurs alternatifs (Holmer, 2013 ; Lindekilde, 2012b ; Mastroe et Szmania, 2016). Le résultat est une diversité d'approches et de méthodologies d'évaluation qui abordent également une diversité d'angles, souvent de manière pragmatique, des programmes visés.

g. Les difficultés concernant la politisation du phénomène

La prévention de l'extrémisme violent ne se fait pas dans n'importe quel contexte. Il y a un sentiment d'urgence concernant un phénomène que l'on pense être en augmentation et qui est alimenté par la pression des médias et une inquiétude dans la société. Les répercussions de ce contexte particulier se font sentir dans les évaluations, notamment en raison des pressions politiques qui se traduisent par un impératif d'agir, la démonstration de l'efficacité des mesures et la perception de la fiabilité de la démarche. Des chercheuses et chercheurs ont déjà fait part de leurs préoccupations quant à la fiabilité des données communiquées concernant certains programmes de « déradicalisation » soumis à d'énormes pressions politiques (Horgan et Braddock, 2010 ; Mastroe et Szmania, 2016). Hirschi et Widmer expliquent mieux le défi des personnes évaluatrices dans ce contexte :

Les incidents qui ont un lien avec l'extrême droite ont souvent une résonance publique considérable, ce qui entraîne une émotivité du phénomène lui-même. L'évaluation se trouve alors confrontée au défi de se positionner de telle sorte qu'elle puisse paraître tout aussi digne de confiance à ceux qui participent à une mesure qu'au cercle plus large des personnes (par exemple en politique ou dans les médias) qui sont intéressées ou touchées par celle-ci (2012, p. 172).

Par conséquent, de nombreux programmes ont été politiquement motivés et non fondés sur des preuves. Cela se traduit le plus souvent par des objectifs irréalistes et peu clairs ou par des théories du changement non testées ou trop ambitieuses qui posent des problèmes évidents pour le travail d'évaluation (Baruch et al., 2018 ; Ris et Ernstorfer, 2017). Le centre de prévention et d'insertion à la citoyenneté (CPIC), à Pontourny en France, en est un exemple (Albert et al., 2020). Conçu en urgence pour détourner les jeunes du djihad, il a été l'un des échecs de la stratégie nationale française les plus médiatisés de ce type de mesures. Malgré cette pression politique et médiatique entourant l'évaluation, les personnes intervenantes et chercheuses s'entendent pour dire qu'il est nécessaire de garantir la volonté politique d'entreprendre des évaluations indépendantes, scientifiquement fondées et d'en tirer des enseignements (Chowdhury Fink et al., 2013).

1.3 QUE DISENT LES PRÉCÉDENTES RECENSIONS DE LA LITTÉRATURE SUR LES PROGRAMMES ÉVALUÉS ET LEURS MÉTHODOLOGIES D'ÉVALUATION ?

Avant de commencer la présentation des résultats de notre étude, il est certainement pertinent de faire un bref détour sur les différentes tentatives passées de répertorier plus ou moins systématiquement les études évaluatives et d'identifier les forces et les faiblesses de ces recensions afin de comprendre la pertinence d'un nouvel effort dans cette direction. Dans cette section, nous présentons un aperçu historique de ces revues et terminons par une brève conclusion générale sur celles-ci et sur les éléments qu'une nouvelle recension systématique peut apporter au domaine de l'évaluation de ce type de programmes. À titre d'illustration, le Tableau 1 montre le nombre d'études qui ont été retenues dans ces revues et également dans notre étude ainsi que les raisons de l'exclusion de certaines de notre étude.

Lum, Kennedy et Sherley (2006) ont été l'une des premières équipes à réaliser une revue systématique sur les programmes de contre-terrorisme. Ils notent que la plupart des actions n'ont jamais été évaluées, ce qui signifie l'absence de base factuelle pour ces programmes. En fait, d'une base de données originale de 20 000 titres, ils n'ont retenu que 7 programmes de contre-terrorisme qui ont fait l'objet d'une évaluation satisfaisante (Tableau 1).

Tableau 1. Revues systématiques et recensions de la littérature de programmes de prévention de la radicalisation et de l'extrémisme violents

Revue systématique ou recension de la littérature	Retenues par cette étude	Études exclues			
		CT*	SDP*	NE*	F*
Bellasio et al., 2018	28/48	7	3	2	8
Carthy, Doody, Cox, O'Hora, et Sarma, 2020	0/14	14			
Feddes et Gallucci, 2015	11/55	6	19	2	17
Gielen, 2017	25/73	4	38	3	3
Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021	47/51	2	1		1
Madriaza, Ponsot, Marion, Monnier, Ghanem et al., 2017; Madriaza et Ponsot, 2015	12/23	6	3		2
Mastroe et Szmania, 2016	16/43	7	14	1	5
Pistone et al., 2019	17/38	5	12	2	2
Pratchett, Thorp, Wingfield, Lowndes et Jabbar, 2010	1/18	4	6		7
Taylor et Soni, 2017	1/7	5		1	

CT : Étude classifiée comme mesure antiterroriste, non directement liée à la prévention ou qui ne porte sur aucun programme spécifique.

SDP : Étude sans données primaires ou avec des données anecdotiques.

NE : Étude non évaluative.

F : Publications inaccessibles ou fusionnées avec une autre publication qui a utilisé le même échantillon et analyse.

Pratchett et ses collègues (2010) ont réalisé la première recension (*rapid evidence assessment*) de programmes de prévention évalués, notamment au Royaume-Uni. Ils ont répertorié 18 études, dont la première date de 1996, bien avant le tournant préventif⁹. À la différence des revues subséquentes, la recension de l'équipe de Pratchett a eu accès à des évaluations sous-embargo de la part du *Department for Communities and Local Government* du Royaume-Uni. D'après les auteures et auteurs, aucune des 18 sources documentaires n'a permis de savoir quelles mesures ont été les plus efficaces pour changer l'attitude des communautés à l'égard de l'extrémisme violent, et ce, en raison de la qualité méthodologique des études qui avaient un design fondamentalement qualitatif. Deux ans plus tard, Christmann et al. (2012) n'ont trouvé que deux programmes additionnels au Royaume-Uni.

En 2014, le projet Impact Europe a publié un rapport qui résumait trois études, dont les deux dernières abordaient le sujet de la prévention (van Hemert et al., 2014). L'une était basée sur un échantillon non représentatif de 100 programmes de prévention trouvés sur deux bases de données entre les années 2000 et 2014. Le but n'a pas été de répertorier les programmes évalués, mais

de décrire une série de variables présentes dans les publications. Cependant, ils ont identifié 52 programmes qui affirmaient avoir effectué des évaluations. Celles-ci ont été, pour la plupart, des évaluations d'impact, suivies par des évaluations de processus et des évaluations économiques. Toutefois, la plupart de ces évaluations ne sont pas allées au-delà d'un simple retour des commentaires des personnes participantes concernant le programme.

L'autre étude, en revanche, était spécifiquement axée sur les méthodes d'évaluation et a servi de base à l'étude de Feddes et Gallucci (2015). À partir d'une recension plus systématique, ces auteures et auteurs ont réussi à identifier 55 publications en lien avec l'évaluation de programme qui ont fait appel à 135 échantillons sur 9 bases de données entre les années 1990 et 2014. Sur les 135 échantillons identifiés, seuls 16 provenaient de programmes qui ont fait l'objet d'une évaluation empirique qui a utilisé des données primaires. Le reste concernait des évaluations anecdotiques¹⁰ (49 %) ou théoriques¹¹ (39 %). La plupart des évaluations identifiées par Feddes et Gallucci (2015) étaient une combinaison d'évaluations d'impact et de mécanismes¹² (46 %), suivies

⁹ C'est l'étude de Knox et Hughes (1996) concernant des programmes communautaires qui visaient à consolider la paix dans la période post-conflit en Irlande du Nord. Cette étude n'a pas été retenue dans notre revue systématique.

¹⁰ Les publications qui décrivaient seulement le programme.

¹¹ Les publications qui ont testé une théorie au moyen d'une recension de la littérature sans avoir recours à des données qualitatives ou quantitatives.

¹² Des évaluations qui expliquent pourquoi un programme a fonctionné (Feddes et Gallucci, 2015).

par des évaluations de processus¹³ (23 %) et économiques (19 %). Bien que la plupart des programmes aient été pensés pour produire des effets à long terme, la grande majorité des évaluations de ces programmes s'est avérée des études transversales. En fait, seuls trois échantillons provenaient d'études quasi expérimentales. Un autre élément qui se répète dans la littérature est l'absence de théories du changement dans les programmes évalués. Dans le cas de cette étude, une théorie du changement a été identifiée dans 12 % des échantillons et, dans la grande majorité des cas (60 %), les évaluations n'étaient basées sur aucune théorie spécifique. L'étude de Feddes et Gallucci (2015), la première dans son genre, est une base importante pour l'avancement des méthodologies d'évaluation dans ce domaine.

En 2016, Mastroe et Szmania ont publié un rapport dont l'objectif était d'examiner la littérature sur les mesures d'évaluation utilisées dans les études empiriques portant sur les programmes de lutte contre l'extrémisme violent entre 2005 et 2016. Ces chercheuses en ont trouvé 43. Malgré la prétention empirique de cette revue, les auteures n'ont identifié que 5 études touchant la prévention et 17 ayant trait au désengagement et à la déradicalisation, qui ont utilisé des données primaires. Le reste de ces 43 études touchait notamment les activités mises en place afin d'atteindre les objectifs des programmes évalués (*Output evaluation*). La plupart de ces études ont été des recherches descriptives et, comme dans le cas de recensions passées, aucune de ces études n'a utilisé de designs expérimentaux. Les auteures concluent que, lors de la publication de leur rapport, il n'y avait guère de consensus concernant l'efficacité des programmes de prévention en raison du manque d'études empiriques et elles remettent en question la confiance que l'on peut accorder aux résultats d'une recherche qui ne rend pas transparente dans ses publications la méthodologie utilisée pour parvenir à ces résultats.

En 2017, sur la base d'une autre revue systématique, Madriaza et al. ont identifié 23 études concernant la prévention : 14 de ces études faisaient preuve de transparence méthodologique, ce qui impliquait une description complète des personnes participantes, des outils de cueillette et d'analyse de données et de la méthodologie. Également en 2017, Gielen a publié une revue de la littérature dans laquelle elle a identifié 73 études, dont 14 étaient des évaluations de résultats, 14, des évaluations de processus, et le reste, des évaluations fondées sur la théorie, c'est-à-dire qu'elles s'appuyaient sur la littérature du domaine pour évaluer un programme. Cette revue constate également une augmentation des évaluations à partir de 2016, ce qui a été corroboré dans d'autres revues récentes (Bellasio et al., 2018). Gielen

mentionne que, parmi les évaluations de résultats ou d'effets, seules 4 études étaient quantitatives, dont deux ont utilisé un design quasi expérimental, 9 étaient des études qualitatives et une seule étude a utilisé une méthode mixte.

L'étude de Bellasio et al. (2018) est probablement l'une des plus détaillées dans le domaine de l'évaluation de programmes de prévention et s'inscrit dans la même direction que celles de Feddes et Gallucci (2015) et Mastroe et Szmania (2016) concernant l'examen spécifique des méthodes d'évaluation. À partir d'une revue systématique conduite entre 2013 et 2018¹⁴, ces chercheuses et chercheurs identifient 48 études en contreterrorisme et en prévention de l'extrémisme violent. La grande majorité de ces études portait sur des évaluations d'impact, suivies par des évaluations de processus. Vingt-quatre d'entre elles ont utilisé des méthodes qualitatives et, à la différence d'autres recensions évoquées, ces personnes chercheuses ont trouvé un nombre plus important d'études qui utilisent des méthodes mixtes (14). En revanche, l'absence d'études longitudinales semble être une tendance persistante dans ce domaine. Bellasio et ses collègues (2018), par exemple, notent que la plupart des études ont été transversales et/ou intermédiaires et seules six études comprenaient des mesures ex post. L'évaluation de la qualité méthodologique est moins développée dans cette étude. Cependant, il existe quelques éléments qui peuvent être mentionnés. En fait, une vaste majorité des études identifiées a été réalisée par des évaluatrices et évaluateurs externes, ce qui est un indicateur d'indépendance de l'évaluation, et plus de la moitié a fait l'objet d'un processus d'évaluation aveugle par des pairs ou est passée par l'examen d'un comité indépendant. Un résultat intéressant concerne l'approche d'évaluation. Bellasio et ses collègues n'ont pas pu identifier d'approche théorique claire dans 33 de ces études.

En 2018 et 2019, d'autres recensions ont été réalisées qui ne se sont toutefois pas focalisées spécifiquement sur les méthodologies d'évaluation des études. La revue de Pistone et ses collègues (2019) a répertorié 112 publications faites entre 1989 et 2017. Seules 38 d'entre elles présentaient des résultats sur l'évaluation et 15 utilisaient des données primaires, le reste étant des réflexions quant à l'efficacité des mesures mises en place. En 2020, Carthy et ses collègues (2020) ont publié une revue systématique portant sur les mesures en matière de contre-discours en prévention de l'extrémisme violent, dans laquelle ils ont retenu 14 études. C'est la première revue systématique à effectuer une méta-analyse des effets de ce type de mesures. Étonnamment, en dépit du titre de la revue, aucune de ces études ne visait

¹³ C'est l'évaluation sur la façon dont le programme a été mis en œuvre (Feddes et Gallucci, 2015).

¹⁴ Onze de ces études ont été recommandées par des personnes expertes et publiées avant 2013.

spécifiquement l'extrémisme violent, même si ce sont des programmes qui peuvent s'appliquer à ce contexte. Les équipes de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021) ont réalisé récemment deux revues systématiques basées sur un modèle de santé publique concernant les programmes de prévention primaire et secondaire, dans le cas de la première, et de prévention tertiaire, dans le cas de la deuxième. Ces équipes ont retenu respectivement 33 et 18 études dans leurs revues. À l'exception de Carthy et al. (2020), ces revues sont les premières à utiliser un outil spécifique d'évaluation de la qualité méthodologique comme critère d'inclusion et à considérer des études ayant exclusivement analysé des données primaires, ce qui peut être considéré comme une amélioration dans la qualité de ces études. En revanche, une limite importante de cette étude et de plusieurs autres présentées précédemment est associée au biais de publication. Une grande partie des études recensées concerne, en fait, seulement des publications dans des revues spécialisées. Par ailleurs, dans le cas de la première revue systématique, l'équipe de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021) a pu identifier une évolution par rapport aux recensions précédentes quant au nombre d'études qui utilisent des méthodologies mixtes (16), laquelle a été une recommandation dans la littérature spécialisée (Williams et Kleinman, 2014). Dix études ont été exclusivement qualitatives et seules 7 ont utilisé des méthodes quantitatives. Quant à la qualité méthodologique, les études retenues par cette équipe ont reçu, en moyenne, 6,7 sur une échelle de 0-10, les études avec un score de 3 ou moins ayant été exclues¹⁵. Finalement, Zeuthen (2021) a récemment publié une nouvelle revue systématique sur les programmes de prévention tertiaire où l'auteure a trouvé 15 études qui répondaient à ses critères d'inclusion, en plus de 34 autres publications consistant principalement en des recensions de la littérature dans ce domaine ou dans des domaines connexes. L'étude de Zeuthen mentionne avoir procédé à une évaluation de la qualité des études, mais malheureusement, elle ne décrit pas clairement la méthode d'évaluation utilisée.

Le nombre de recensions plus ou moins systématiques de programmes évalués dans ce domaine a permis de dresser un état des lieux de l'évaluation ainsi que des défis auxquels ce type de recension peut être confronté et des lacunes existantes en matière de connaissances qui peuvent être abordées dans cette nouvelle étude. Malgré le nombre important d'études antérieures, elles présentent des limites et des biais évidents qui justifient un nouvel effort dans ce domaine.

La première limite est sans aucun doute le pragmatisme avec lequel ces revues ont été abordées, qui est en partie dû à la nature embryonnaire de l'évaluation en PEV. En 2013, Chowdhury Fink et ses collègues publient un rapport basé sur un atelier organisé à Ottawa avec des personnes chercheuses et praticiennes. Ces derniers s'entendent sur un point : les difficultés pour évaluer ce type de programmes sont évidentes et plusieurs actrices et acteurs ont souligné l'importance d'avoir une approche pragmatique en matière d'évaluation. Ces propos ont également été soulignés par d'autres chercheuses et chercheurs (Marret et al., 2017; Romaniuk, 2015). Romaniuk le résume de cette façon : « C'est surtout le pragmatisme qui semble prévaloir, avec des évaluateurs qui recueillent les données qu'ils peuvent avec les ressources disponibles » (2015, p. 36). Cette conclusion peut également s'appliquer au cas des recensions de la littérature présentées dans cette dernière section. Le faible nombre d'évaluations empiriques et le besoin de déterminer l'efficacité de ces mesures de prévention ont poussé les chercheuses et chercheurs à inclure dans leurs analyses des publications concernant l'évaluation de programmes sans avoir recours à des données primaires ou reposant sur des données anecdotiques. Une part importante de ces recensions inclut, par exemple, des évaluations sans analyse de données. La remarque de Mastroe et Szmania (2016) concernant la fiabilité des données communiquées, dans le cas des programmes de « déradicalisation », peut également être une source de préoccupation dans les cas de publications portant sur des évaluations d'extrait (*output evaluations*) et dans le cas des études qui utilisent des données anecdotiques. Les dernières études sont toutefois plus prometteuses en matière de matériel empirique et montrent une amélioration dans le domaine. Elles ont permis, en effet, de tirer des conclusions plus spécifiques en ce qui concerne l'efficacité des mesures mises en place.

Un deuxième élément à prendre en compte est le biais en faveur d'une évaluation de l'efficacité de ces programmes au détriment d'une évaluation plus compréhensive. Tout au long de cette introduction, et en particulier dans cette dernière section, la question prédominante a porté principalement sur ce que nous savons des effets de ces programmes et très peu sur la manière dont ces programmes ont été appliqués, dans quels contextes ils sont les plus efficaces et quels mécanismes sous-jacents au fonctionnement de ces programmes peuvent être mobilisés pour améliorer leur performance. Parmi les recensions identifiées, seule Gielen (2017) a tenté d'aller au-delà de cette question en utilisant l'approche « réaliste » de l'évaluation. Cet accent a donné lieu à une quête incessante de méthodes quantitatives rigoureuses

¹⁵ Cette analyse a été réalisée par nous sur la base des résultats présentés dans les rapports des équipes d'Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021)

en tant qu'indicateur d'une bonne performance d'évaluation, qui puisse principalement répondre à la question de l'efficacité. C'est le cas de l'utilisation de l'échelle des méthodes scientifiques du Maryland par Bellasio et ses collègues (2018) comme outil d'évaluation de la qualité des études, qui ne concerne que les designs quantitatifs. Ce biais peut être dû au caractère récent de l'évaluation dans ce domaine, une période où la première question à laquelle il faut répondre concerne sans doute l'efficacité de ces programmes. Cependant, comme nous le verrons plus loin dans ce rapport, le domaine est suffisamment mature pour intégrer de nouvelles questions d'évaluation.

Le biais de publication reste un enjeu à traiter. La littérature grise s'est avérée fondamentale dans le domaine et elle a été prise en considération par une grande partie des recensions abordées. Toutefois, comme nous le verrons dans le cas de notre revue systématique, ces recherches n'ont pas été exhaustives. Les actrices et acteurs en dehors du milieu académique ont été beaucoup plus actifs pour combler ces lacunes.

L'évaluation de la qualité méthodologique est également un autre enjeu. Seules les études des équipes de Carthy et al. (2020) et de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021) ont appliqué des outils spécifiques pour évaluer la qualité de la méthode dont les résultats sont utilisés comme critère d'inclusion et d'exclusion. L'étude de l'équipe de Bellasio et al. (2018) a utilisé le Maryland Scientific Methods Scale (MSMS), mais cet outil permet seulement de déterminer la qualité de l'étude en fonction du type de design. De cette façon, une étude randomisée sera par défaut une étude de bonne qualité, indépendamment d'autres facteurs.

Un dernier élément à considérer est la spécificité de ces recensions. À l'exception de l'étude de Carthy et de ses collègues (2020) sur les mesures de contre-discours, ou les recensions concernant les programmes au Royaume-Uni (Christmann et al., 2012; Pratchett et al., 2010; Taylor et Soni, 2017), les recensions ont tendance à aborder la prévention de l'extrémisme violent de façon large, sans prendre en considération la diversité des mesures d'évaluation mises en place et la particularité des contextes d'intervention. Gielen (2017) a identifié au

moins 8 types différents de programmes de prévention. Déjà, en 2011, Neumann affirmait que ces initiatives pouvaient être potentiellement illimitées. Des milieux, tels que l'éducation ou le système correctionnel, et des régions, comme l'Amérique du Nord, l'Afrique, l'Asie du Sud et du Sud-est, ont mis en place des mesures de prévention adaptées aux contextes locaux, qui nécessitent des analyses spécifiques. Une situation similaire existe dans le cas des programmes destinés à l'extrême droite, dont très peu ont été évalués et qui présentent également des défis spécifiques d'évaluation (Bellasio et al., 2018; Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hirschi et Widmer, 2012). Taylor et Soni (2017) ont abordé l'expérience de « *Prevent* » dans le milieu éducationnel anglais, mais d'après nos critères, la plupart des études retenues ne portaient pas sur un programme spécifique. Les revues systématiques des équipes de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021) sur les programmes de prévention primaire, secondaire et tertiaire et l'étude de Zeuthen (2021) sont également un avancement en la matière du fait de leur concentration sur des types spécifiques de prévention. Cependant, le besoin d'évaluer l'efficacité de ces programmes dans des contextes spécifiques demeure un enjeu majeur dans le domaine.

Ces éléments de conclusion nous permettent de justifier la pertinence d'une nouvelle revue systématique sur les méthodes d'évaluation, qui intègre de nouvelles connaissances et de nouvelles façons d'aborder cette question de recherche, notamment en ce qui concerne l'accès à la littérature grise, l'évaluation de la qualité des méthodes choisies et, bien sûr, en ce qui concerne les nouvelles données empiriques et les nouvelles études qui n'ont pas été prises en compte. Bien qu'au moins trois de ces revues (Bellasio et al., 2018; Feddes et Gallucci, 2015; Mastroe et Szmania, 2016) se soient concentrées sur les méthodes d'évaluation, elles présentent néanmoins un grand nombre des lacunes susmentionnées, que nous espérons combler avec cette nouvelle étude.



Méthodologie pour l'élaboration de notre revue systématique

La méthode utilisée pour élaborer notre revue systématique est basée sur les méthodes de revue du Campbell Collaboration¹⁶. Nous définissons une revue systématique comme étant « une revue d'une question clairement formulée qui utilise des méthodes systématiques et explicites pour identifier, sélectionner et évaluer de façon critique les recherches pertinentes, et pour recueillir et analyser les données des études qui sont incluses dans la revue » (Moher et al., 2009, p. 1).

2.1 OBJECTIF

Le but global de cette revue systématique est de répertorier les évaluations de programmes de prévention de l'extrémisme violent (PEV) publiées jusqu'en décembre 2019.

En plus de cet objectif général, quatre objectifs spécifiques sont avancés :

1. Identifier les méthodologies utilisées dans l'évaluation des programmes de prévention de l'extrémisme violent.
2. Identifier les lacunes de la littérature sur l'évaluation de ces programmes.
3. Évaluer la qualité méthodologique des études évaluatives existantes dans le domaine.
4. Formuler des recommandations pour l'évaluation de ces programmes.

¹⁶ Disponible au <https://www.campbellcollaboration.org>. Une version plus complète de notre méthode se trouve dans les annexes.

2.2 QUESTION DE RECHERCHE

Notre question principale est, par conséquent, la suivante : Quelles sont les principales recommandations, en matière d'évaluation de programmes de prévention dans le domaine de l'extrémisme violent, que l'on peut tirer de la littérature ?

2.3 CRITÈRES D'INCLUSION

Nous avons ciblé toutes les études basées sur des données probantes primaires¹⁷ qui ont évalué des programmes de PEV, publiées jusqu'en décembre 2019. Ces programmes ont eu pour but de réduire ou de supprimer les conditions de risque qui peuvent rendre un individu ou un groupe plus vulnérable à l'extrémisme violent ou qui le poussent vers la récidive¹⁸. Nous avons inclus toutes les études qui ont eu pour but d'apprécier ou de juger un programme, un projet ou une stratégie de PEV, y compris les études qui ne mentionnent pas directement le mot « évaluation ». La population cible de ces études n'a pas été considérée comme un critère d'inclusion. Par conséquent, nous ciblons toutes les évaluations des programmes de PEV primaire, secondaire et tertiaire¹⁹ qui essaient de modifier les attitudes, les émotions ou les comportements de ces individus ou de ces groupes, des familles ou de l'entourage de ces individus et des personnes intervenantes travaillant dans ce domaine. En revanche, les évaluations de programmes qui prennent en charge les victimes directes ou indirectes²⁰, les évaluations des mesures de contreterrorisme ainsi que les études qui évaluent les stratégies continentales ou portent un jugement global sur une approche continentale sont exclues.

Par conséquent, l'unité d'analyse est constituée par les études publiées et non par les publications, puisqu'une même publication peut faire état de plus d'une étude. Nous avons considéré qu'il y avait plus d'une étude dans une publication lorsqu'il y avait plus d'un échantillon qui était analysé indépendamment et qui présentait des résultats indépendants.

Les études devaient être rédigées en anglais, en français ou en espagnol (langues parlées par les membres de l'équipe de recherche).

2.4 VARIABLES CODÉES

Chaque document scientifique a été codé en fonction d'une grille de codage global composée de nombreuses dimensions principales²¹ :

- Description générale des études retenues
- Identification des auteures et auteurs des études retenues
- Niveau de prévention
- Type d'extrémisme
- Type de modèle d'évaluation
- Type d'évaluateur ou d'évaluatrice
- Design ou approche méthodologique en fonction de l'approche globale
 - de la manipulation de variables des participantes et participants
 - du nombre d'observations
 - des temps d'observations
 - du nombre de variables indépendantes
 - du nombre de variables dépendantes
- Outils de collecte de données
- Type d'action ciblée
- Échantillon
- Population cible
- Milieu cible
- Type d'indicateurs utilisés ou résultats obtenus
- Types d'effets
- Limites de la recherche

Nous avons utilisé le *Mixed Methods Appraisal tool* (MMAT) (Hong, Pluye, Fàbregues, Bartlett, Boardman et al., 2018 ; Hong et Pluye, 2019) comme outil d'évaluation de la qualité méthodologique des documents scientifiques retenus dans cette revue systématique. À la différence d'autres outils d'évaluation, le MMAT permet, au sein du même instrument, d'évaluer des études qualitatives, quantitatives descriptives, expérimentales, quasi expérimentales et mixtes, ce qui représente la variété d'études que nous nous attendions à retrouver lors de la recherche. Le MMAT se compose de 25 variables réparties en 5 groupes représentant les cinq types d'études mentionnés ci-dessus. Chaque étude peut obtenir un score entre 0 et 5.

¹⁷ Les données secondaires, pour leur part, désignent les données collectées par une personne autre que la chercheuse, le chercheur ou son équipe. Les exemples de sources de données secondaires dans les sciences sociales sont les recensements, les données collectées par les ministères, les registres organisationnels et d'autres données collectées qui avaient à l'origine d'autres objectifs de recherche.

¹⁸ Voir les définitions clés utilisées dans les annexes.

¹⁹ Idem.

²⁰ Les familles des individus ayant entamé ce processus peuvent être considérées comme victimes indirectes des groupes extrémistes. Cependant, ici, nous comprenons les « victimes » comme les personnes et leurs familles qui ont été la cible d'attentats, de tentatives d'attentat ou d'actions violentes de la part de groupes extrémistes.

²¹ La liste complète et les définitions des dimensions et des variables incluses sont présentées en annexe.

2.5 STRATÉGIE DE RECHERCHE DE LA REVUE DE LITTÉRATURE

À partir des critères d'inclusion et d'exclusion et des mots clés identifiés, nous avons utilisé trois stratégies de recherche :

- Revue de la littérature scientifique
- Revue de la littérature grise
- Autres sources

2.5.1 Littérature scientifique

Nous avons fait appel à un bibliothécaire expert dans le domaine des sciences sociales et humaines qui, sur la base de nos critères, a réalisé une recherche dans 21 bases de données. Ces bases de données nous ont permis de repérer non seulement des articles scientifiques ou thèses, mais aussi un nombre important de littérature grise et de communications lors de congrès. Nous avons aussi obtenu l'accès à la base de données d'une revue systématique du Réseau des Praticiens Canadiens pour la Prévention de la Radicalisation et de l'Extrémisme Violent (RPC-PREV) (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021). Les deux bases de données ont été fusionnées.

2.5.2 Littérature grise

Afin de réduire le « biais de publication » (Rothstein, Sutton et Borenstein, 2005), nous avons effectué une recherche approfondie de la littérature grise en utilisant Google. En outre, 228 sites Web des organisations travaillant dans le domaine de la radicalisation et de la lutte contre l'extrémisme violents ont été parcourus manuellement pour répertorier d'autres documents. Ces organismes ont été sélectionnés à partir d'une cartographie de pôles d'expertises en PEV réalisée par la Chaire UNESCO-PREV de façon préalable (<https://chaireunesco-prev.ca/lereseau/cartographie/>). D'autres organismes ont été inclus au cours de la recherche.

2.5.3 Autres sources

En plus des documents identifiés à l'aide de la stratégie de recherche décrite ci-dessus, nous avons comparé nos résultats avec des revues ou recensions fréquemment citées dans la littérature (Tableau 1). Nous avons également consulté par courriel électronique 14 personnes expertes pour vérifier si elles connaissaient d'autres études en prévention de l'extrémisme violent.

2.6 PROCÉDURE

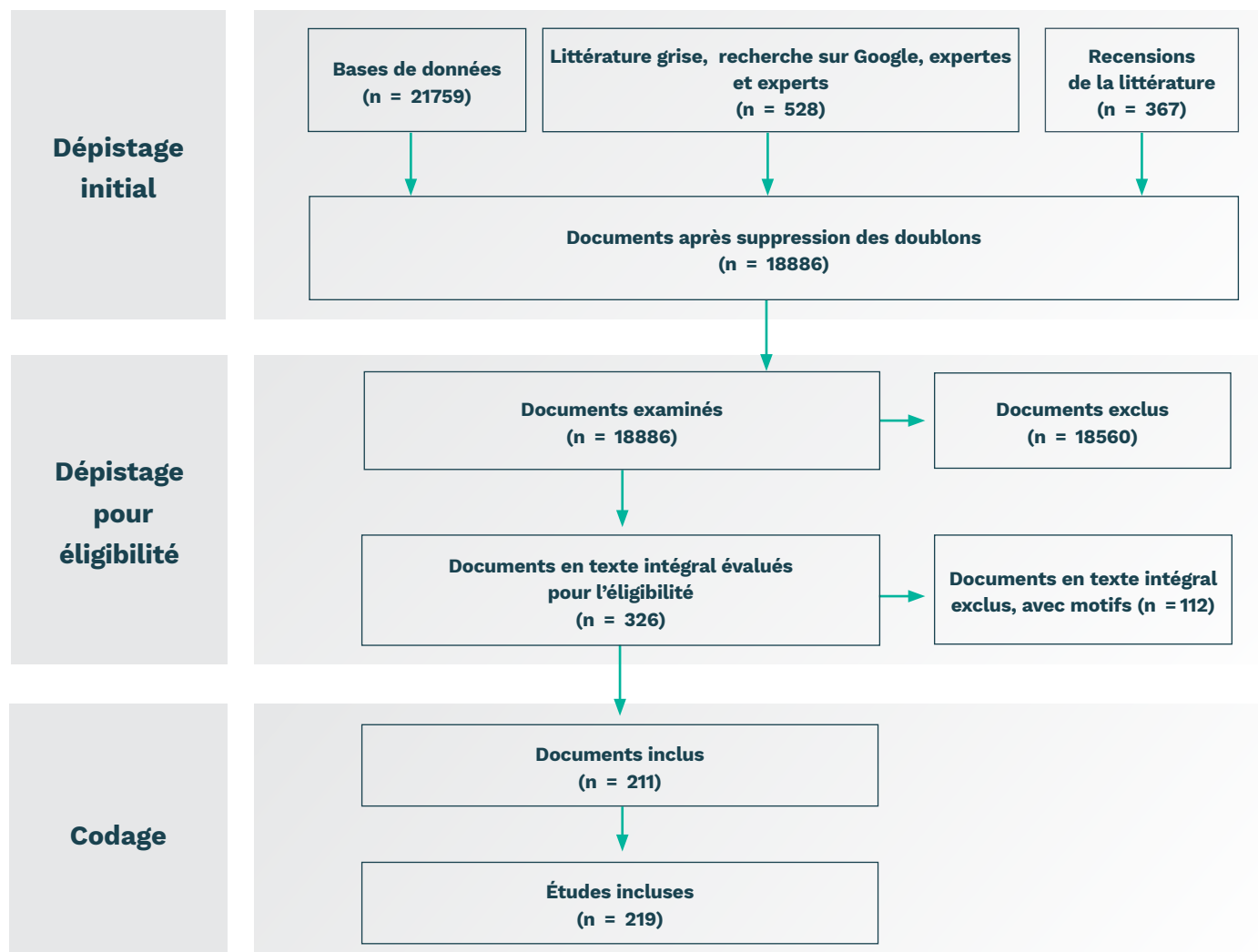
Avant le début de nos travaux, il y a eu une période de formation pour les cinq auxiliaires retenus pour clarifier les concepts et la méthodologie de travail. Nous avons utilisé deux sources de données bibliographiques. L'une provient de la revue systématique faite par l'équipe du RPC-PREV (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021), qui comporte certains mots clés en commun. Cette base de données couvrirait toutes les publications existantes jusqu'en janvier 2018. Une nouvelle revue bibliographique a été conduite en utilisant les critères déjà mentionnés. Pendant que le bibliothécaire faisait cette cueillette de documents scientifiques, les auxiliaires de recherche ont révisé la littérature grise disponible sur les sites web des organismes mentionnés ci-dessus. Dès que la cueillette de données de la littérature grise a été achevée, les bases de données ont été fusionnées et les doublons ont été éliminés. Les 14 personnes expertes ont été contactées pendant cette période.

Pour éliminer les études non admissibles, les auxiliaires de recherche et le chercheur principal ont passé au crible les titres et les résumés des documents identifiés lors de la recherche documentaire. Pendant cette première phase, l'ensemble de l'équipe a codé les premiers 700 documents afin d'assurer une certaine cohérence. Les désaccords ont été analysés et résolus. Cette phase a aussi servi de formation à l'équipe. Par la suite, deux codeurs ont révisé chaque document. Pour s'assurer que l'accord entre les personnes évaluateurs était adéquat, un Kappa de Cohen a été calculé. Pendant ce codage initial, nous avons travaillé de façon itérative, c'est-à-dire que chaque binôme d'évaluateurs a travaillé sur un nombre limité d'articles. Ensuite, le Kappa de Cohen a été calculé : si le résultat était de moins de 0,6 (le minimum admissible), le binôme révisait leurs désaccords; si le résultat du Kappa était au-dessus de 0,6, le binôme continuait à coder la tranche suivante de documents. Le Kappa final a été de 0,86.

Le nombre total de publications sélectionnées était de 211, mais certaines publications faisaient état de plus d'une étude. En conséquence, le nombre total d'études retenues a été de 219. Nous avons considéré qu'il y avait plus d'une étude dans une publication lorsqu'il y avait plus d'un échantillon qui était analysé indépendamment.

Nous avons utilisé le modèle PRISMA (<http://www.prisma-statement.org>) pour consigner les résultats des recherches documentaires dans un organigramme (Figure 1).

Figure 1. Flux décisionnel Prisma



2.7 LIMITES DE CETTE RECHERCHE

Cette revue systématique présente des limites dans la collecte et l'analyse des données qui doivent être prises en compte. La première limite est un biais de publication concernant les langues utilisées. Une partie importante des programmes évalués, en particulier les programmes de prévention de l'extrémisme violent de droite, ont été rédigés dans des pays où sont parlées des langues qui n'ont pas été prises en compte dans cette étude. C'est notamment le cas pour les études faites en allemand, en néerlandais et celles écrites dans les langues nordiques. Nous n'avons donc pas eu accès à un nombre significatif de ces études, ce qui biaise particulièrement nos résultats en ce qui concerne ce type d'extrémisme.

Une autre limite importante est associée à l'évaluation de la qualité méthodologique. D'une part, l'évaluation de la qualité méthodologique dépendait principalement des informations que nous avons pu obtenir des publications. Il est possible que de nombreuses publications aient

été limitées dans l'espace, en particulier la littérature académique, en raison des limites des revues scientifiques et que beaucoup d'informations n'aient pu être incluses, ce qui limite notre capacité d'analyse. D'autre part, il est également recommandé que la première étape de l'évaluation de chaque étude, ainsi que l'évaluation de la qualité des méthodes, soit réalisée par deux personnes à travers un jugement en aveugle entre les deux. Cependant, principalement en raison du grand nombre de publications retenues et des contraintes de temps et de ressources de notre équipe, l'évaluation de la qualité des méthodes, notamment, n'a été effectuée que par une seule personne par étude. En fait, diverses personnes ont évalué et codé différentes études. La première étape de l'évaluation des études à inclure a, en revanche, été réalisée par deux personnes évaluatrices. Malgré la qualité de l'outil d'évaluation et ses critères détaillés, l'évaluation de la qualité des méthodes comporte une composante de subjectivité qui ne peut être négligée.



Résultats

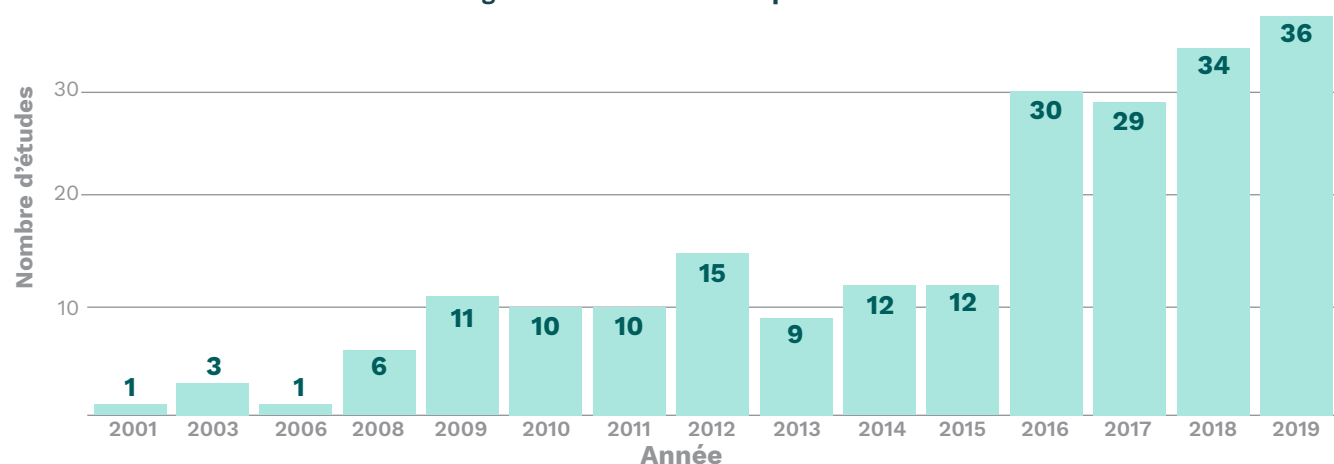
3.1 QUE SAVONS-NOUS GLOBALEMENT DES PROGRAMMES ÉVALUÉS ET DE LEURS ÉTUDES ?

La Figure 1 de la section précédente illustre le flux décisionnel Prisma concernant les articles identifiés et retenus. Sur une base de données originale de 18 886 documents scientifiques, nous avons retenu **211 publications et 219 études**. La différence entre le nombre de publications et le nombre d'études est due au fait que treize publications portaient sur plus d'une étude ($n=36$) et que onze études ont été diffusées dans plus d'une publication. C'est le cas, par exemple, du rapport de la *University of Amsterdam* (2013) et de l'article de Feddes et Gallucci (2015) qui se sont basés sur la même étude²². Les études étant l'unité d'analyse de cette recherche, nous nous concentrerons désormais sur la description des 219 études répertoriées.

3.1.1 Quelle est l'évolution des études par année ?

Tel qu'observé par Gielen (2017) et l'équipe de Bellasio et al. (2018), on note **un point tournant en 2016 concernant le nombre d'études évaluatives publiées**, à partir duquel s'amorce une tendance croissante. Si, pendant les années précédentes, la moyenne d'études publiées est de neuf par année, elle grimpe à 30 études à partir de 2016 (Figure 2). Malgré les difficultés pour mettre en place des évaluations dans le domaine, ce constat atteste des efforts réels entrepris afin de combler cette lacune. Une culture d'évaluation semble commencer à s'installer parmi les actrices et acteurs de la prévention de l'extrémisme violent.

Figure 2. Nombre d'études par année



²² L'information a été fusionnée et seul le document le plus complet a été considéré dans le nombre final d'études. Lorsque des contradictions apparaissaient entre les deux documents, nous avons priorisé l'information du document le plus complet.

3.1.2 D'où viennent ces études ? (nombre d'études par pays et continent)

La plupart des recensions existantes permettent de constater les efforts consentis par les pays européens en matière d'évaluation, une tendance que confirme cette revue systématique (Figure 3). Près de la moitié des évaluations ont porté sur des programmes européens, dont 22,8% ont été réalisées au Royaume-Uni qui a été le précurseur de ce type d'initiatives. Les Pays-Bas (n = 12), l'Allemagne (n = 5), le Danemark (n = 5), la France (n = 4) et la Suisse (n = 4)²³ ont également entrepris des évaluations de différents programmes. Toutefois, compte tenu de nos critères d'inclusion, il convient de noter que le nombre d'études pourrait être beaucoup plus élevé dans les pays qui publient dans des langues autres que l'anglais, le français et l'espagnol.

Les études évaluatives ne se concentrent que sur quelques pays de la planète.

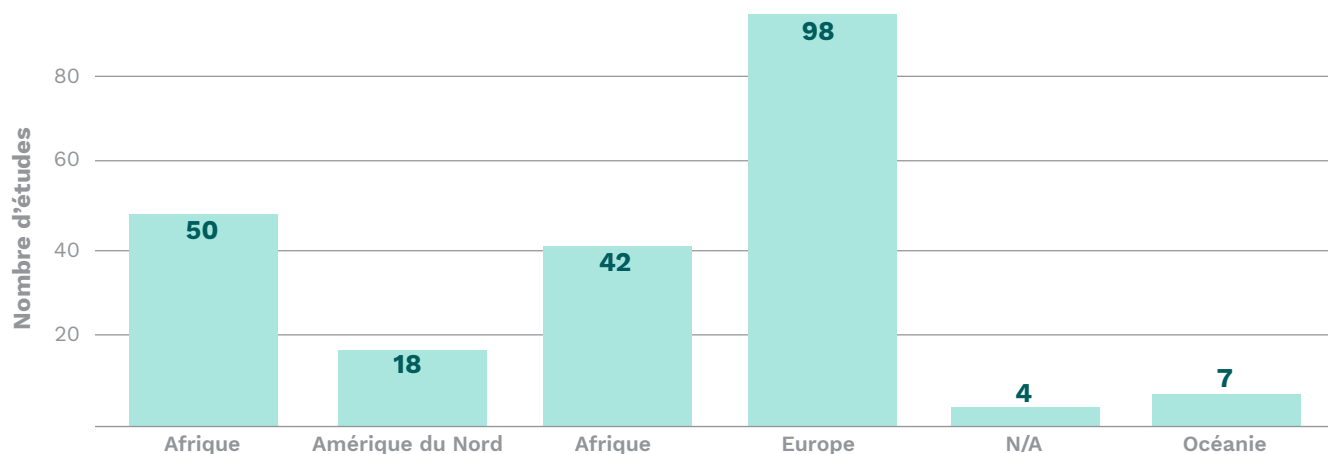
Outre l'importance de l'Europe en matière d'évaluation, cette revue systématique a identifié un nombre important d'études sur d'autres continents concernés par l'extrémisme violent, qui n'ont pas été répertoriées dans les recensions passées. C'est le cas, notamment, de l'Afrique (n = 50) et de l'Asie (n = 42). En Afrique, le Kenya (n = 9), la Somalie (n = 6), le Niger (n = 5), le Nigéria (n = 4) sont les pays ayant fait l'objet du plus grand nombre d'évaluations sur le continent. En Asie, l'Indonésie (n = 16) est particulièrement concernée ainsi que le Pakistan (n = 9) et, dans une moindre mesure, les Philippines (n = 4). Fait notable, l'Indonésie est le deuxième pays au monde qui a le plus grand nombre d'évaluations identifiées par cette revue systématique après le Royaume-Uni.

Un autre élément à souligner est le fait que 31 sur 50 études identifiées en Afrique et 29 sur 42 en Asie ont

été rédigées en anglais malgré le fait que la langue officielle du pays en question était autre. En Afrique, alors que près de la moitié des études ont évalué des programmes dans des pays où le français est une langue officielle ou une langue d'usage, aucune d'entre elles n'a été rédigée en français. Le fait que ces études soient rédigées en anglais présente un avantage évident pour la diffusion potentielle des résultats à plus grande échelle et les échanges scientifiques. Toutefois, cela constitue, en même temps, un obstacle pour de nombreux acteurs, actrices et intervenantes, intervenants sur le terrain qui ne peuvent pas nécessairement y accéder. Ce constat pourrait indiquer que les évaluations n'ont pas été pensées pour permettre une rétroaction vis-à-vis les personnes qui conçoivent et mettent concrètement en œuvre les programmes sur le terrain et, ainsi, pour avoir un impact sur l'amélioration des pratiques, notamment en ce qui concerne les pays non occidentaux, mais plutôt pour répondre au besoin des bailleurs de fonds. Cependant, ce constat n'est qu'une hypothèse parmi d'autres, car nous n'avons pas les moyens de corroborer si d'autres moyens de mobilisation des connaissances ont été utilisés en retour de ces programmes.

L'Amérique du Nord, qui a une longue tradition d'évaluation dans des domaines connexes, est étonnamment l'une de celles où il y a le moins d'évaluations, même si on note une augmentation significative à partir de 2016. Les États-Unis ont toutefois consenti certains efforts à ce sujet avec un certain nombre d'études (n = 15). De son côté, jusqu'en 2019, le Canada ne comptait que trois programmes évalués ayant fait l'objet d'une publication. Toutefois, l'étude de terrain réalisée par le RPC-PREV concernant les programmes de prévention secondaire et tertiaire au Canada indique que, parmi les 26 programmes répertoriés dans ce pays en prévention de l'extrémisme violent, au moins cinq ont ou vont entamer un processus d'évaluation (Hassan, Ousman et al., 2020) (Tableau 2).

Figure 3. Nombre d'études par continent²⁴



²³ Cette information est basée sur une seule publication qui a été divisée en 7 études différentes, dont 4 retenues pour notre étude.

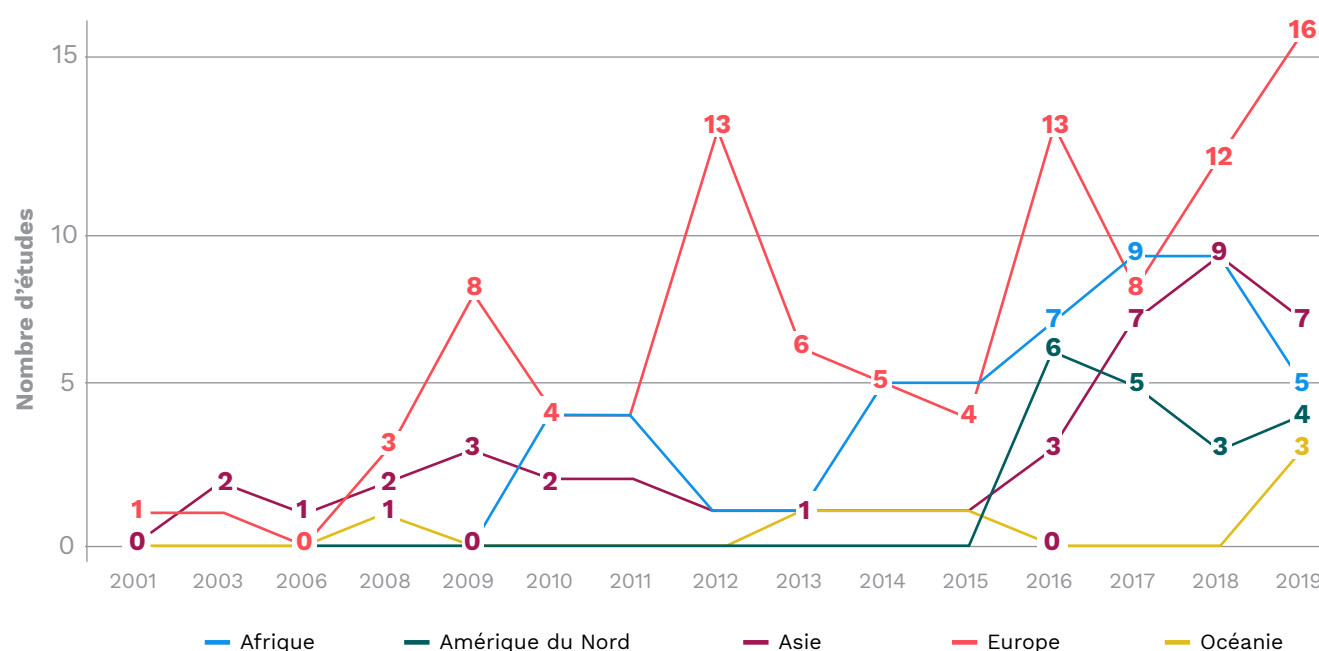
²⁴ N/A concerne les programmes, notamment en ligne, qui ne ciblaient pas une région ou un pays spécifique. Pour une description détaillée des études des programmes en ligne, voir la section concernant les études de cas à la [page 85](#).

Tableau 2. Les quinze pays ayant fait le plus d'évaluations (nombre d'études)

Pays	n	%	Pays	n	%	Pays	n	%
Royaume-Uni ²⁵	50	22,6	Pakistan	9	4,1	Niger	5	2,3
Indonésie	16	7,2	Australie	7	3,2	France	4	1,8
États-Unis	15	6,8	Somalie	6	2,7	Nigéria	4	1,8
Pays-Bas	12	5,4	Allemagne	5	2,3	Philippines	4	1,8
Kenya	9	4,1	Danemark	5	2,3	Suisse	4	1,8

L'évolution du nombre d'études par année et par continent peut expliquer en partie la montée des études à partir de 2016. L'Europe a eu une évolution par pics au cours des deux dernières décennies et l'Afrique, une tendance croissante à partir de 2014. Pour leur part, l'Asie et l'Amérique du Nord montrent une augmentation considérable à partir de l'année 2016 (Figure 4).

Figure 4. Évolution des études par année et par continent



3.1.3 Littérature académique ou littérature grise ?

La majorité des études retenues par cette revue systématique a été diffusée dans des publications non académiques. D'un point de vue global, cela démontre que l'intérêt à l'évaluation ne concerne pas exclusivement le milieu académique et qu'un nombre important d'organismes non universitaires y sont impliqués (Tableau 3). En revanche, la littérature académique a connu une augmentation linéaire depuis 2016 (Figure 5) qui peut s'expliquer pour le grandissant intérêt des approches quantitatives et designs expérimentaux et quasi expérimentaux des dernières années (voir les sections 3.3.3 et 3.3.4).

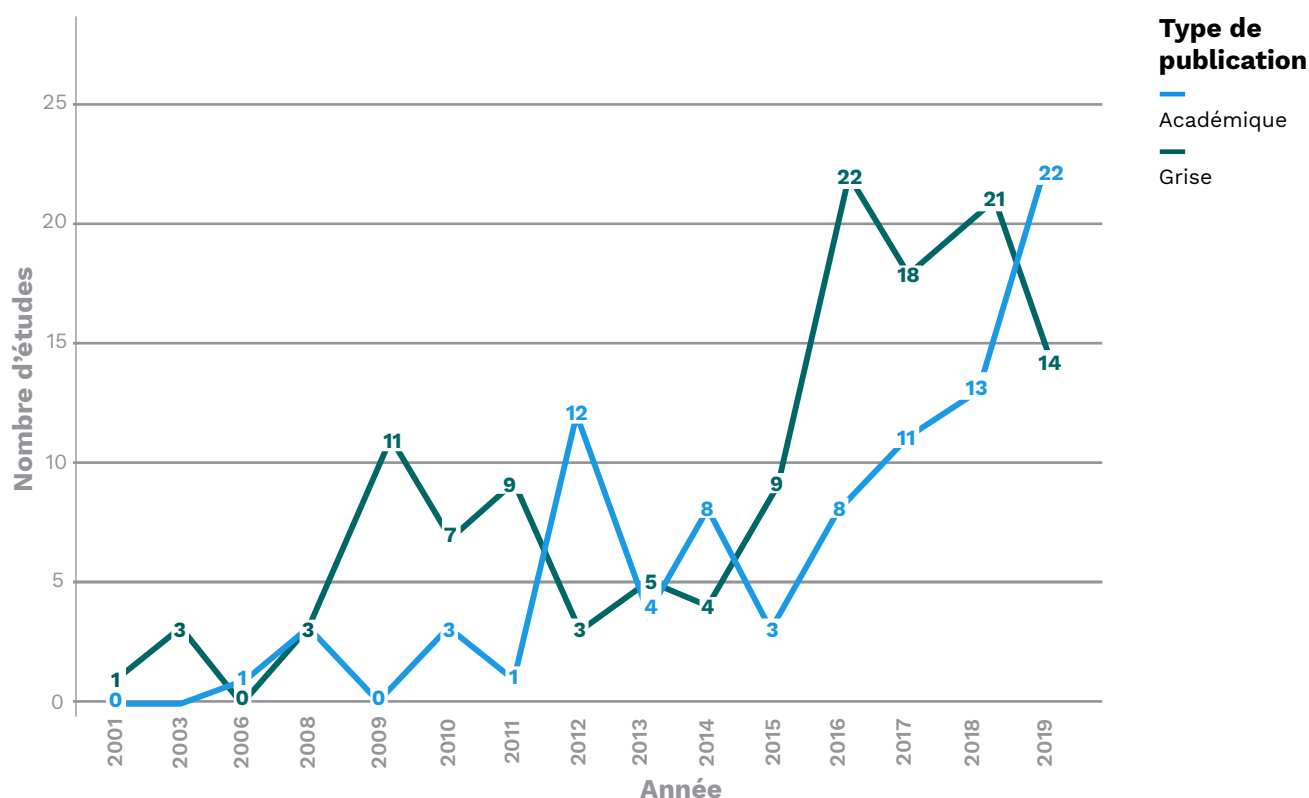
²⁵ Le Royaume-Uni comprend l'Angleterre, l'Écosse, l'Irlande du Nord et le Pays de Galles. Certaines études n'indiquent le « Royaume-Uni » que comme origine géographique du programme, tandis que d'autres l'indiquent clairement. Afin de limiter la confusion qui pourrait être causée si nous montrions dans le même tableau les catégories « Royaume-Uni » et « Angleterre », par exemple, nous avons étiqueté toutes les études provenant de ces pays comme « Royaume-Uni ».

Tableau 3. Type de publication par continent (nombre d'études)²⁶

Continent	Type de publication		Total
	Académique	Grise	
TOTAL	89	130	219
Afrique	22 % (11)	78 % (39)	100,0 % (50)
Amérique du Nord	27,8 % (5)	72,2 % (13)	100,0 % (18)
Asie	42,9 % (18)	57,1 % (24)	100,0 % (43)
Europe	47,5 % (47)	52 % (51)	100,0 % (99)
N/A	25 % (1)	75 % (3)	100,0 % (4)
Océanie	100,0 % (7)	0,0 % (0)	100,0 % (7)

Cependant, la situation n'est pas la même selon les régions. Dans le cas de l'Afrique et de l'Amérique du Nord, cette majorité est écrasante : les trois quarts des publications proviennent de la littérature grise. C'est le cas, notamment, des évaluations réalisées pour le programme de développement américain (USAID) ou faites par des organismes tels que « Search for common ground ». En revanche, toutes les études menées en Océanie (Australie) ont été publiées dans des revues scientifiques traditionnelles, tandis qu'en Europe et en Asie, la situation est davantage équilibrée entre les deux types de publications.

Figure 5. Type de publication par année



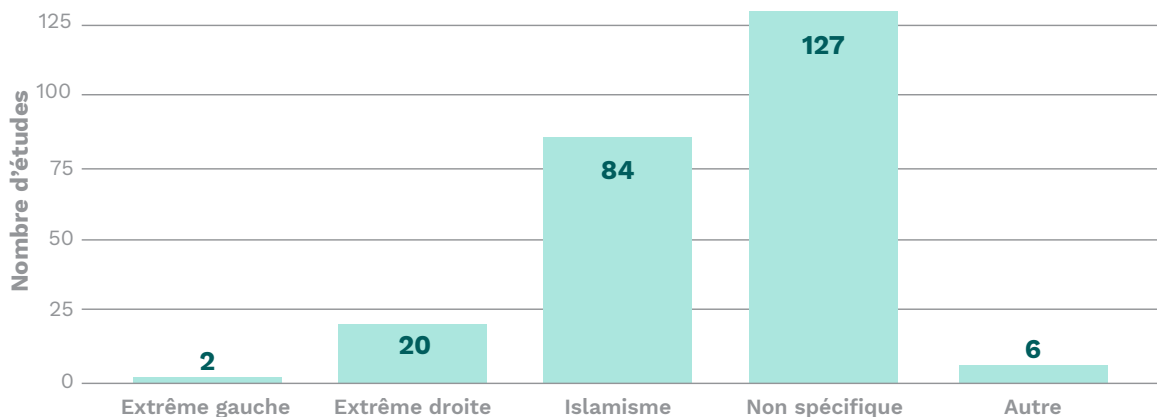
²⁶ Sauf indication contraire, les pourcentages des tableaux sont calculés par ligne. Les études recensées pouvant présenter parfois plusieurs ou aucun des éléments décrits, le total par ligne n'est pas toujours égal à 100%.

3.1.4 Quel type d'extrémisme les programmes évalués ciblent-ils ?

La majorité des programmes évalués (n=127) ciblent la radicalisation et l'extrémisme violents en général sans prendre en compte un type d'extrémisme particulier.

Viennent ensuite ceux qui ciblent officiellement l'extrémisme violent en lien avec l'islamisme (n = 84) et, enfin, l'extrême droite (n = 20).

Figure 6. Type d'extrémisme ciblé²⁷

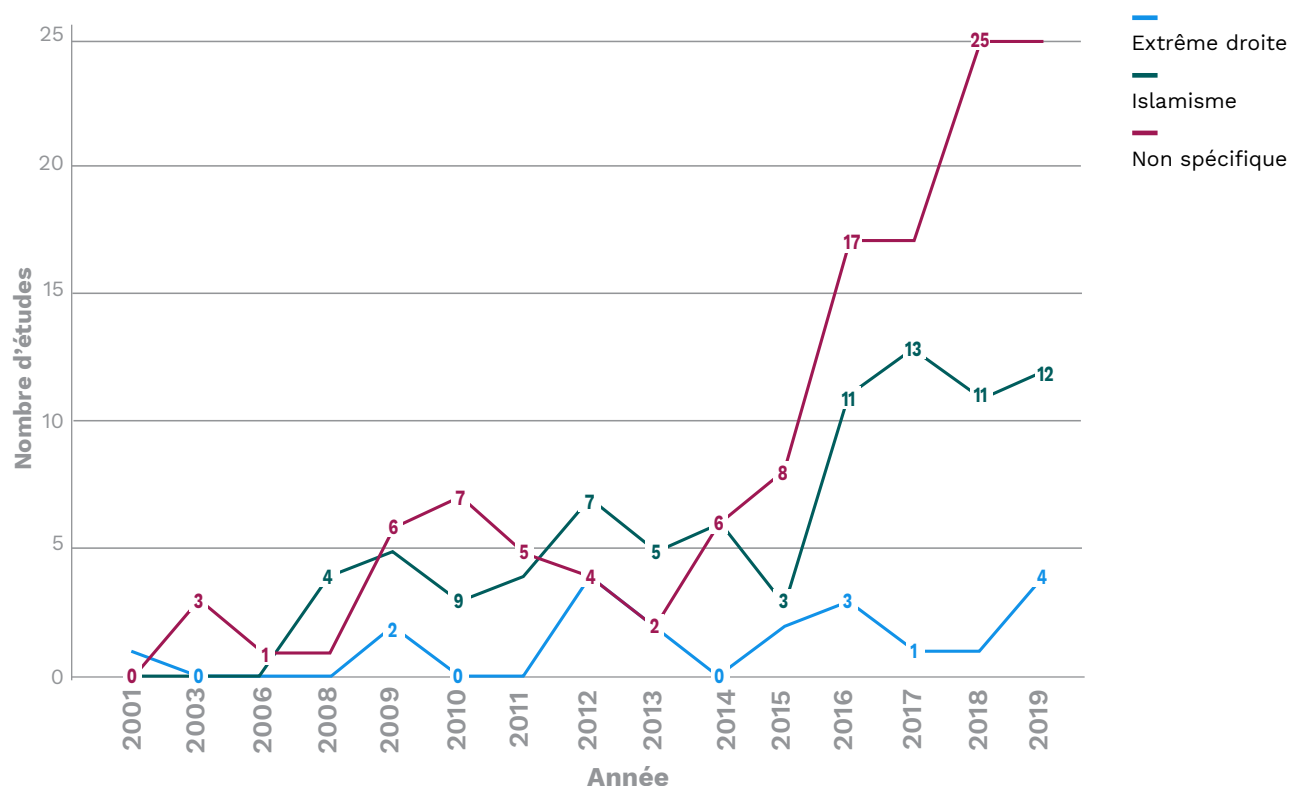


Les programmes évalués non spécifiques ont connu une nette augmentation à partir de 2016 (Figure 7). Alors qu'ils ne représentaient en moyenne que quatre programmes évalués annuellement jusque-là, ils représentent depuis lors, en moyenne, vingt-deux programmes par an. Par rapport à l'ensemble des évaluations, on retrouve ce type de programmes, notamment en Afrique (68 % du total des programmes du continent), en Amérique du Nord (61,1%) et en Europe (60,2%). L'évolution des publications liées à ce type de programme évalué se distingue selon les continents. En Europe, on observe une tendance croissante à partir de 2014 et une augmentation considérable en 2018 (n = 12) et en 2019 (n=13). En Afrique, on observe une légère augmentation à partir de 2015 (n = 4) qui demeure stable dans les dernières années. En Amérique du Nord et en Asie, on constate une légère augmentation à partir de 2016 et 2017 respectivement (n=3). L'essor des programmes non spécifiques évalués s'explique, notamment, par une certaine prise de conscience quant aux effets de la fausse association entre Islam et terrorisme et le potentiel stigmatisant de ce type de programme, notamment pour la communauté musulmane. En Europe, l'évaluation négative de la stratégie nationale du Royaume-Uni, « *Prevent* », en est l'une des manifestations les plus connues qui a contribué à ce virage conceptuel en termes d'approche. Le premier « *Prevent* » (2007-2011) visait directement la communauté musulmane et a été considéré comme un facteur important de stigmatisation de cette communauté dans le pays (Busher et al., 2019 ; Kundnani, 2012 ; Romaniuk, 2015). Comme suite à ces évaluations, cette stratégie nationale a été modifiée en 2011 pour élargir son approche

à toutes les formes d'extrémisme (Busher et al. 2019). Plusieurs pays, dont le Canada, ont pris note de cette évaluation et ont préféré miser sur des approches plus généralistes d'intervention. En Afrique et en Asie, plusieurs intervenantes et intervenants sont également réticents face à l'utilisation de concepts tels que la radicalisation ou la prévention de l'extrémisme violent dit « islamiste » pour des raisons assez semblables, auxquels s'ajoutent toutefois des contextes particulièrement sensibles dus à la place de la religion musulmane dans certaines de ces sociétés (Madriaza et al., 2017). L'utilisation de tels concepts peut ainsi briser le lien de confiance avec la communauté et empêcher le travail d'intervention (Madriaza et al., 2017). Dans ces régions, la prévention de l'extrémisme violent passe souvent par des initiatives d'insertion professionnelle et éducationnelle, autrement dit, de la prévention primaire. Dès lors, plusieurs personnes intervenantes considèrent comme peu utile ou pertinente l'utilisation de ces concepts (Madriaza et al., 2017). Nonobstant l'importance de ce changement d'approche quant à la façon d'aborder la prévention de l'extrémisme violent et son évaluation, il n'en reste pas moins que, dans les faits, ces programmes ciblent la plupart du temps l'extrémisme islamiste. En effet, hormis dans les pays occidentaux et en Amérique latine où l'extrémisme de droite et de gauche est aussi présent, nombre de pays concernés par cette catégorie généraliste, notamment en Afrique, sont aux prises avec ce seul type d'extrémisme. En d'autres termes, si l'on s'en tient aux cibles effectives des programmes de prévention et *a fortiori* de l'évaluation, la très grande majorité des études portant sur l'évaluation concerne encore aujourd'hui l'extrémisme dit « djihadiste » ou « islamiste ».

²⁷ Nous avons codé cette catégorie à partir de l'information présente dans les publications. Nous avons codé un type d'extrémisme en particulier quand il a été explicitement indiqué dans le texte. Dans les cas où le texte ne mentionnait aucun type d'extrémisme particulier ou indiquait qu'il visait la radicalisation ou l'extrémisme en général, nous avons codé « Non spécifique ». L'addition peut faire plus de 219 études puisque plusieurs programmes ciblent plus d'un type d'extrémisme.

Figure 7. Nombre d'études par type d'extrémisme ciblé et par année



Ces données peuvent être confirmées en analysant l'évolution des études évaluatives de programmes de prévention de l'extrémisme violent dit « islamiste ». En Europe, le nombre d'évaluations de programmes qui ciblent directement l'extrémisme islamiste est effectivement décroissant depuis 2016. En Afrique et en Asie, en revanche, la courbe est relativement stable au cours des dernières années et demeure un pourcentage mineur par rapport à l'ensemble des études.

Seconde remarque, on note une augmentation relative du nombre d'études ciblant l'extrême droite, surtout à partir de 2015 (Figure 7), dont huit sur dix proviennent d'Europe. Cela est d'autant plus important que les recensions passées notaient le faible nombre d'études évaluatives répertoriées qui ciblaient l'extrême droite et

le considéraient comme une lacune importante dans le champ (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021; Hirschi et Widmer, 2012; Widmer et al., 2007).

Troisième remarque, les évaluations mentionnant spécifiquement l'extrême gauche sont quasiment inexistantes, les deux qui sont identifiées proviennent de l'Europe.

Enfin, dernier élément, alors que la question des réseaux sociaux est omniprésente dans le débat public, on note qu'il y a très peu de programmes de prévention en ligne qui ont fait l'objet d'évaluations (16 programmes au total, voir section 3.6.2).

3.1.5 À quel niveau de prévention les programmes évalués interviennent-ils ?

Sur la base d'un modèle de santé publique et selon la catégorisation utilisée par Hassan et son équipe (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021), nous avons fait la différence entre cinq types de programmes (Figure 8) :

Les programmes de prévention primaire (n = 48), à savoir les programmes de prévention universelle qui ciblent la population générale non identifiée comme étant à risque ;

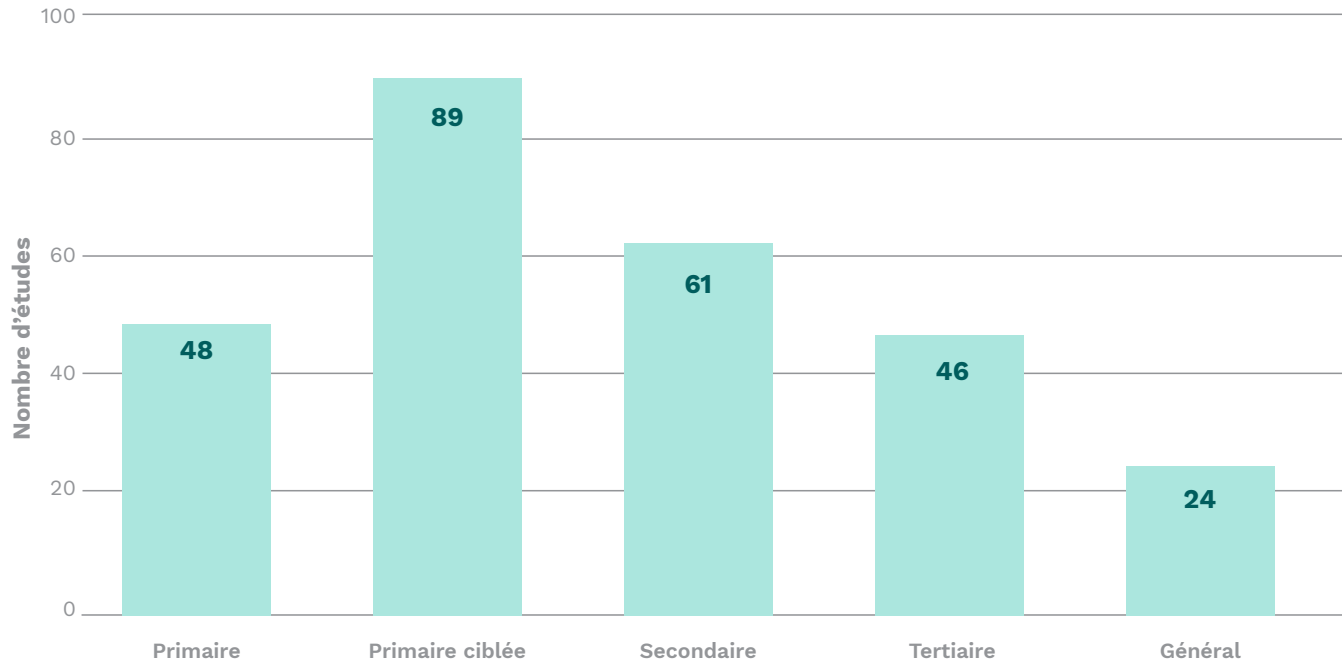
Les programmes de prévention primaire ciblée (n = 89) qui, bien qu'universels, ciblent une communauté spécifique ou un secteur de la population considéré à risque (par exemple, les jeunes ou les communautés musulmanes) ;

Les programmes de prévention secondaire (n = 61) qui ciblent les individus ou les groupes considérés à risque ou qui se retrouvent dans les étapes initiales du processus de radicalisation et de l'extrémisme violents ;

Les programmes de prévention tertiaire (n=46) qui ciblent les individus ou les groupes qui sont déjà engagés dans le processus de radicalisation et de l'extrémisme violent dans les phases finales, qui appartiennent à un groupe extrémiste ou qui ont commis des actes associés à l'extrémisme violent ;

Les programmes de prévention générale (n= 24) sont ceux dont les documents scientifiques n'ont pas spécifié si les programmes utilisaient l'un des types de prévention mentionnés.

Figure 8. Nombre d'études par type de prévention du programme évalué

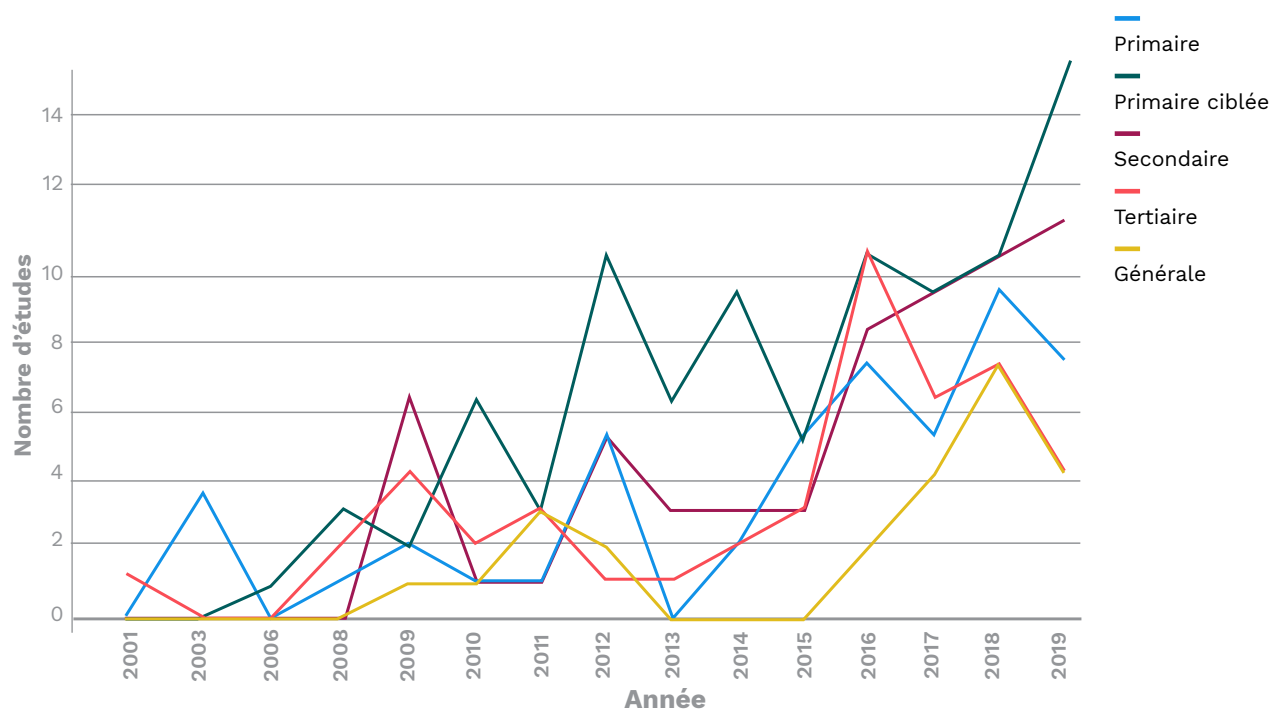


De manière générale et de façon logique, l'évolution annuelle des programmes évalués par niveau de prévention (Figure 9) suit la même tendance que l'on observe pour l'ensemble des programmes évalués, à savoir une augmentation continue depuis 2016. On note toutefois quelques nuances pour certains niveaux de prévention. La tendance à la croissance est plus claire dans les cas des programmes de prévention primaire (sauf la dernière année), primaire ciblée et secondaire. Cela peut s'expliquer, notamment, dans le cas des deux premiers, en partie pour les mêmes raisons qui s'appliquent à l'évolution des différents types d'extrémisme ciblé, à savoir le virage en faveur d'une approche de prévention plutôt globale que spécifique. En revanche, les programmes évalués en prévention tertiaire ont connu une nette diminution depuis 2016 qui contraste avec la hausse des évaluations pour les autres niveaux de prévention. Il s'agit souvent, pour la plupart, de programmes déployés en milieu correctionnel ou dans le système de probation.

L'une des difficultés pour évaluer ce type de programmes, mentionnées dans la partie précédente du rapport, réside dans le faible nombre de cas et dans la difficulté d'avoir accès aux données et aux personnes participantes dans ces milieux, ce qui pourrait expliquer cette diminution. Cela demeure toutefois seulement une hypothèse qui peut sembler d'ailleurs d'autant plus surprenante que cette diminution apparaît dans une période marquée par le retour progressif des personnes revenant de la zone de conflit syrienne. Il faut rappeler que ces données concernent les programmes évalués et non les programmes mis en place. Par conséquent, nous observons une tendance décroissante à évaluer ce type de programme plutôt qu'une diminution de programmes de prévention tertiaire qui, du reste, pourraient faire l'objet d'évaluations ultérieures.

La majorité des programmes évalués sont ainsi les programmes de prévention primaire et primaire ciblée.

Figure 9. Programmes par niveau de prévention par année



Le Tableau 4 présente les résultats concernant le niveau de prévention par continent et par type d'extrémisme. La répartition par continent montre l'emphase mise sur un type de stratégie ou un autre. Les programmes évalués en prévention primaire ciblée, par exemple, sont en proportion plus grande dans les régions occidentales (Océanie [Australie] = 66,7 % Europe = 53,1 % ; Amérique du Nord = 44,4 %) que dans d'autres régions. En Europe, ils sont suivis par les programmes de prévention secondaire (32,7 %) et tertiaire (26,5 %). En d'autres termes, les programmes européens évalués ont ainsi tendance à se focaliser sur la population considérée à risque. Cela pourrait être une conséquence de l'interprétation des facteurs expliquant le phénomène de radicalisation et d'extrémisme violents qui prédomine dans ces programmes européens, c'est-à-dire une interprétation basée moins sur les facteurs sociaux que dans d'autres parties du monde et plus sur les facteurs individuels.

Tableau 4. Nombre d'études par niveau de prévention par continent et par type d'extrémisme²⁸

		Primaire		Primaire ciblée		Secondaire		Tertiaire		Générale	
		n	%	n	%	n	%	n	%	n	%
Total		48	17,8 %	89	33 %	61	22,6 %	48	17,8 %	24	8,9 %
Continent	Afrique	20	40,0 %	12	24,0 %	13	26,0 %	5	10,0 %	8	16,0 %
	Amérique du Nord	4	22,2 %	8	44,4 %	3	16,7 %	1	5,6 %	5	27,8 %
	Asie	6	14,3 %	13	31 %	9	21,4 %	11	26,2 %	5	11,9 %
	Europe	17	17,3 %	52	53,1 %	32	32,7 %	27	26,5 %	6	6,1 %
	N/A	1	25,0 %	0	0,0 %	2	50,0 %	2	50,0 %	0	0,0 %
	Océanie	0	0,0 %	4	66,7 %	2	33,3 %	1	16,7 %	0	0,0 %
Type d'extrémisme	Extrême droite	3	15,0 %	5	25,0 %	13	65,0 %	8	40,0 %	0	0,0 %
	Islamisme	11	13,1 %	40	47,6 %	25	29,8 %	23	27,4 %	6	7,1 %
	Non spécifique	34	27 %	48	38,1 %	32	25,4 %	18	14,3 %	20	15,9 %

²⁸ L'addition du nombre total d'études peut représenter plus de 219 études parce que certaines études ont été codées sur plus d'une catégorie. Pour la même raison, la somme des pourcentages des études par type d'extrémisme et par continent peut donner des résultats supérieurs à 100 %.

En Afrique, en revanche, ce sont les programmes évalués de prévention primaire qui apparaissent comme plus nombreux dans la région. Comme nous l'avons évoqué, d'autres études ont montré que la prévention sur ce continent passe souvent par des initiatives d'insertion professionnelle et éducative, en réponse au recrutement et à l'embrigadement dans les groupes extrémistes qui exploitent les problèmes socio-économiques pour offrir aux jeunes des alternatives (Madriaza et al., 2017). Cependant, ces résultats montrent également que l'extrémisme violent est perçu comme un problème qui cible l'ensemble de la société et qu'il doit être abordé de façon universelle. Il ne faut pas oublier qu'une partie importante de ces initiatives évaluées a été financée et mise en place par des organismes occidentaux et pourrait correspondre davantage à l'interprétation occidentale du phénomène en Afrique.

Les programmes évalués ciblant tous les types d'extrémisme utilisent les niveaux de prévention les plus universels : prévention primaire ciblée (38,1 %) et primaire (27 %). En revanche, les programmes évalués en prévention de l'extrême droite ne semblent pas toucher la prévention primaire et universelle, mais se concentrer davantage en prévention secondaire et tertiaire. Cette observation est cohérente avec celle d'autres revues systématiques qui ont dénombré très

peu de programmes en prévention primaire et secondaire en lien avec l'extrémisme de droite (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021). Ce sont notamment les programmes de prévention tertiaire qui ont été identifiés par la littérature dans le cas de l'extrême droite, et ce, en raison de l'importance des programmes « Exit », particulièrement en Europe (Björger et Horgan, 2009). Les programmes de prévention de l'extrémisme violent dit « islamiste » abordent plutôt la question selon une approche de prévention primaire ciblée, portant notamment sur les communautés musulmanes.

3.1.6 Quelles sont les actions ciblées ?

Parmi les études repérées, nous avons identifié des évaluations de trois types d'action ciblée : les évaluations d'une stratégie nationale, les évaluations d'une partie des actions mises en place dans le cadre d'une stratégie nationale (les évaluations partielles) et les évaluations d'un projet ou d'un programme spécifique (Tableau 5). Sans surprise, l'évaluation d'un projet est la plus fréquemment réalisée et représente près de 61,8 % des évaluations tandis que les évaluations partielles d'une stratégie nationale représentent 29,1 % et les évaluations d'une stratégie nationale au complet ne représentent que 9,1 % du total d'études.

Tableau 5. Type d'action ciblée par continent, par niveau de prévention et par type d'extrémisme

		Stratégie nationale		Stratégie partielle		Projet	
		n	%	n	%	n	%
Total		20	9,1 %	64	29,1 %	136	61,8 %
Continent	Afrique	2	4,0 %	11	22,0 %	37	74,0 %
	Amérique du Nord	2	11,1 %	4	22,2 %	12	66,7 %
	Asie	1	2,4 %	11	26,2 %	30	71,4 %
	Europe	15	15,5 %	34	35,1 %	48	49,5 %
	N/A	0	0,0 %	0	0,0 %	4	100,0 %
	Océanie	0	0,0 %	2	28,6 %	5	71,4 %
Niveau de prévention	Primaire	4	8,3 %	13	27,1 %	31	64,6 %
	Primaire ciblée	12	13,5 %	25	28,1 %	52	58,4 %
	Secondaire	6	10,0 %	16	26,7 %	38	63,3 %
	Tertiaire	5	11,1 %	13	28,9 %	27	57,4 %
	Générale	5	20,8 %	7	29,2 %	12	50,0 %
Type d'extrémisme	Extrême droite	1	5,0 %	4	20,0 %	15	75,0 %
	Islamisme	10	12,0 %	30	36,1 %	43	51,8 %
	Non spécifique	13	10,3 %	30	23,8 %	83	65,9 %

Les projets évalués, largement majoritaires, présentent les pourcentages les plus élevés par continent, par niveau de prévention et par type d'extrémisme. Toutefois,

cette supériorité n'est pas constante dans toutes les catégories. Par exemple, elle n'atteint pas 50 % de toutes les actions évaluées en Europe, une région où l'évaluation

d'une partie (35,1%) ou de la totalité des stratégies nationales (15,5%) est très présente. Ceci est dû au nombre d'évaluations qui ont ciblé la stratégie nationale britannique « Prevent » et qui expliquent dans ce continent une proportion relativement plus équilibrée entre les différentes actions ciblées que les autres continents. Sur d'autres continents, les évaluations des politiques publiques (partielles et totales) représentent un tiers de toutes les actions évaluées, ce qui peut témoigner d'un certain progrès du besoin de données probantes sur les efforts nationaux de coordination de la prévention dans le monde.

Les projets évalués sont également moins présents lorsqu'ils travaillent de façon générale sur la prévention et sur l'extrémisme islamiste et sont largement représentés dans le cas de programmes de prévention de l'extrême droite. Nous trouvons le résultat inverse lorsqu'il s'agit d'évaluer les stratégies nationales, lesquelles sont moins présentes dans le cas de la prévention de l'extrême droite et plus présentes dans le cas des évaluations des actions qui travaillent sur tous les niveaux de prévention.

Ce n'est pas un portrait de tous les projets et politiques déployés dans le monde, mais c'est un portrait de l'importance accordée à la quête de preuves. Dans ce cas,

cette quête est moins présente lorsqu'il s'agit d'évaluer les politiques publiques visant à prévenir l'extrémisme de droite et plus lorsqu'il s'agit d'évaluer les stratégies qui fonctionnent à tous les niveaux pour prévenir l'extrémisme islamiste. Cela peut s'expliquer par le fait que les politiques publiques qui se sont concentrées sur la prévention de l'extrémisme violent dit « islamiste » ont fait l'objet d'un débat approfondi sur les éventuels effets iatrogènes de ces actions. En revanche, la prévention de l'extrémisme violent de droite semble être un intérêt récent des politiques publiques.

3.1.7 Le financement a-t-il été rapporté ?

Le Tableau 6 affiche davantage d'informations sur les études où les chercheuses et chercheurs mentionnent la source de financement de l'évaluation. Ceci nous semble être un critère essentiel de transparence dans la mesure où il permet d'identifier les conflits d'intérêts potentiels ainsi que les enjeux éthiques auxquels les personnes évaluatrices sont confrontées. **La source de financement est mentionnée dans moins de la moitié des études**, soit dans une proportion de 43 %, avec les études concernant l'Afrique (52 %) et l'Amérique du Nord (50 %) qui les mentionnent le plus. À l'opposé, une telle mention ne concerne que 35,7 % des études asiatiques.

Tableau 6. Nombre d'études rapportant leur financement par continent, par niveau de prévention et par type d'extrémisme

		Financement rapporté	
		n	%
Total		95	43 %
Continent	Afrique	26	52,0 %
	Amérique du Nord	9	50,0 %
	Asie	15	35,7 %
	Europe	40	41,2 %
	N/A	2	50,0 %
	Océanie	3	42,9 %
Niveau de prévention	Primaire	27	56,3 %
	Primaire ciblée	40	45,5 %
	Secondaire	24	39,3 %
	Tertiaire	16	34,8 %
	Générale	15	62,5 %
Type d'extrémisme	Extrême droite	10	50,0 %
	Islamisme	27	32,5 %
	Non spécifique	59	46,5 %

Plus le programme est spécifique, moins la source de financement est déclarée et, inversement, plus le programme est universel sur le plan de la prévention, plus ce financement est déclaré. Le niveau de prévention

générale est ainsi celui qui détient la plus grande proportion d'études où la source de financement est mentionnée (62,5 %), alors que le pourcentage le plus bas appartient au niveau de prévention tertiaire (34,8 %).

Par type d'extrémisme, la situation est moins ordonnée. Les études qui évaluent des programmes de prévention d'extrême droite déclarent davantage cette source et celles qui ciblent le moins la prévention de l'extrémisme violent dit « islamiste ».

3.2 QUI SONT LES AUTEURES ET AUTEURS DE CES ÉTUDES ?

Comme nous l'avons évoqué lors de l'introduction de ce rapport, pour de nombreux chercheurs et chercheuses, la question de l'organisation responsable ou de l'auteure ou de l'auteur de l'évaluation est particulièrement importante, notamment par rapport aux particularités du domaine (Horgan et Braddock, 2010 ; Marret et al., 2017 ; Mastroe et Szmania, 2016). C'est pourquoi, parallèlement à la collecte des informations des études, nous avons brossé un portrait des auteures et auteurs des publications sélectionnées (Tableau 7). Nous en avons identifié 389 qui ont contribué aux 211 publications retenues dans notre revue systématique²⁹. La plupart des documents sont signés par plusieurs personnes co-auteurs (62,6 %), plus rarement par une seule personne (29,6 %). Une minorité de documents (7,8 %) n'en mentionne aucune, se contentant alors du nom de l'institution qui a publié l'évaluation en question³⁰. **La majorité des auteures et auteurs (89,5 %) n'ont qu'une publication dans notre base de données portant sur les programmes évalués.** 10,5 % des auteures et auteurs ont deux publications et 3,3 % de ces personnes ont trois publications sur ce sujet précis. Le domaine de l'évaluation des programmes de prévention de l'extrémisme violent semblerait ainsi être peu spécialisé. En 2018, Schuurman avait déjà remarqué le faible degré de spécialisation dans le domaine des études de sécurité en général. Son analyse des publications des 9 revues spécialisées, entre 2007 et 2016, a révélé que seulement 72,8 % des auteures et auteurs avaient fait une publication et 13,4 % en avaient fait deux. Analysées depuis cet angle, les personnes qui ont publié des évaluations identifiées dans notre recherche sont relativement plus spécialisées que celles analysées par Schuurman : 14,7 % des auteures et auteurs dans notre base de données avaient une seule publication dans le domaine de la sécurité en général comme seul auteur et 45,1 % comme co-auteur ou co-auteure. Nous avons cependant eu recours à une base de données plus élargie que celle de Schuurman, ce qui peut expliquer cette différence de la sécurité en général³¹. Le pourcentage reste toutefois très bas et le domaine demeure très peu spécialisé.

Tableau 7. Publications d'auteures et d'auteurs détenant plus de deux publications dans le domaine

Auteure/auteur	n	Pays de provenance
Martin Manby	6	Royaume-Uni
Adrian Cherney	4	Australie
Allard Rienk Feddes	3	Pays-Bas
Anne Speckhard	3	États-Unis
Beza Tesfaye	3	États-Unis
James Khalil	3	Royaume-Uni
Lasse Lindekilde	3	Danemark
Paul Thomas	3	Royaume-Uni
Steven E. Finkel	3	États-Unis
Chris A. Belasco	3	États-Unis
Anne Aly	2	Australie
Anthony Sarota	2	États-Unis
Bart Schuurman	2	Pays-Bas
Bertjan Doosje	2	Pays-Bas
Daniel P Aldrich	2	États-Unis
David Schanzer	2	États-Unis
Elisabeth (Lily) Taylor	2	Australie
Emma Belton	2	Australie
Jean-Camille Kollmorgen	2	États-Unis
Jeffrey Swedberg	2	États-Unis
Joel Busher	2	Royaume-Uni
Michele Grossman	2	Australie
Moli Dow	2	Royaume-Uni
Saul Karnovsky	2	Australie
Tinka Veldhuis	2	Pays-Bas
Tufyal Choudhury	2	Royaume-Uni
Cooper Gatewood	2	Royaume-Uni
Iris Boyer	2	Royaume-Uni
Alex Elwick	2	Royaume-Uni
Lee Jerome	2	Royaume-Uni
Jose Liht	2	Royaume-Uni
Sara Savage	2	Royaume-Uni
Oren Ipp	2	Suède
Ardian Shajkovci	2	États-Unis
Michael Neureiter	2	États-Unis
John McCauley	2	États-Unis
Louis Reynolds	2	Royaume-Uni
Therese O'Toole	2	Royaume-Uni
Daniel Nilsson DeHanas	2	Royaume-Uni
Tariq Modood	2	Royaume-Uni
Elena Savoia	2	États-Unis
Marcia A. Testa	2	États-Unis

²⁹ Nous avons retenu 219 études qui ont été identifiées dans 211 publications.

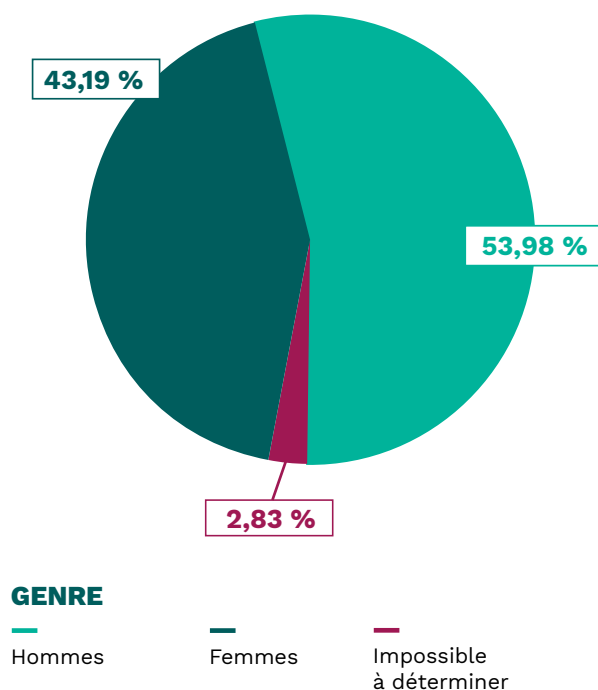
³⁰ C'est le cas, notamment, du rapport de l'« Audit Commission », en 2008, qui a évalué la stratégie nationale « Prevent » du Royaume-Uni. L'« Audit Commission » est un organisme indépendant chargé de veiller à ce que les fonds publics soient dépensés de manière économique, efficiente et efficace au Royaume-Uni.

³¹ Nous avons cherché toutes les publications des auteures et auteurs dans le domaine (radicalisation, extrémisme, contreterrorisme, études de la sécurité, etc.) dans la base de données du système de bibliothèques universitaires de la province de Québec (Sophia) et sur les premières cinq pages de Google Scholar.

3.2.1 Égalité des sexes

Si la majorité des auteures et auteurs sont des hommes (54 %), on note toutefois un nombre assez important de femmes (43,2 %) (Figure 10). Cette zone de parité est toutefois largement en fonction des régions considérées. Par exemple, les auteures et auteurs originaires d'Afrique comptent 78,9 % des hommes auteurs du continent et seulement 21,1 % des femmes alors que ceux et celles originaires d'Asie comptent 52,9 % d'hommes et 36,8 % de femmes. Seule l'Océanie compte plus de femmes (59 %) que d'hommes (41 %).

Figure 10. Pourcentage de femmes et d'hommes ayant publié des études évaluatives



3.2.2 Origine géographique des auteures et auteurs : les États-Unis, le Royaume-Uni et l'Indonésie

Nous avons déjà analysé l'origine géographique des programmes évalués et nous avons remarqué l'importance des régions non occidentales, telles que l'Afrique et l'Asie, ainsi que le faible nombre de programmes évalués en Amérique du Nord, notamment au Canada. Ces données peuvent suggérer que l'évaluation des programmes s'est démocratisée dans le monde au-delà de l'espace occidental. Cependant, l'analyse des auteures et auteurs de publications nous permet de nuancer ces propos. Le Tableau 8 est un résumé des publications des auteures et auteurs par continent. Il inclut le nombre et la moyenne des publications par continent en ce qui concerne les publications retenues dans cette revue systématique ainsi que les autres publications des auteures et auteurs dans le domaine de la sécurité³². Les deux tiers des auteures et auteurs identifiés proviennent de l'Europe (39,1 %) et de l'Amérique du Nord (27,8 %). Les données de l'Europe ne sont pas surprenantes étant donné le nombre d'évaluations réalisées sur le sol européen. Cependant, le cas des auteures et auteurs américains est moins cohérent. L'Amérique du Nord n'ayant réalisé que 18 évaluations, on peut en déduire que la majorité des 108 auteures et auteurs nord-américains, dont 92 provenant des États-Unis, ont été impliqués dans des évaluations à l'extérieur du continent. Le Canada, par exemple, ne compte que 3 évaluations dans notre base de données, cependant, 16 auteures et auteurs canadiens ont été impliqués dans les articles retenus par cette étude. Un cas semblable peut être observé en Océanie. Ne comptant que 7 évaluations réalisées, 27 auteures et auteurs proviennent de ce continent, étant toutes et tous des ressortissants de l'Australie.

Il existe une surreprésentation des personnes évaluatrices nord-américaines et une sous-représentation des personnes évaluatrices africaines

À l'inverse, les 19 auteures et auteurs qui proviennent d'Afrique sont largement moins nombreux que la cinquantaine d'évaluations réalisées sur ce continent, ce qui laisse penser que ces évaluations ont été réalisées en majorité par des personnes évaluatrices provenant de l'extérieur. Le cas de l'Asie est moins dramatique et semble être plus indépendant quant à la réalisation des évaluations : nous avons identifié 68 auteures et auteurs provenant de ce continent et 42 évaluations réalisées.

³² Une définition des critères d'inclusion se trouve dans la note 31.

Tableau 8. Nombre de publications des auteures et auteurs par continent

Continent		Publications comme seul auteur*	Publications comme co-auteur*	Publications dans la base de données**
Total	Moyenne	1,88	3,88	1,14
	Nombre d'auteurs	389	389	389
	Écart type	5,23	8,06	0,48
	Nombre de publications	732	1511	445
Afrique	Moyenne	0,42	0,79	1,00
	Nombre d'auteurs	19	19	19
	Écart type	1,08	0,54	0,00
	Nombre de publications	8	15	19
Amérique du Nord	Moyenne	3,08	4,90	1,17
	Nombre d'auteurs	108	108	108
	Écart type	8,31	12,53	0,46
	Nombre de publications	333	529	126
Asie	Moyenne	1,09	2,51	1,00
	Nombre d'auteurs	68	68	68
	Écart type	3,77	5,00	0,00
	Nombre de publications	74	171	68
Europe	Moyenne	1,70	4,25	1,20
	Nombre d'auteurs	152	152	152
	Écart type	3,15	5,76	0,59
	Nombre de publications	258	646	182
Océanie	Moyenne	2,19	5,07	1,30
	Nombre d'auteurs	27	27	27
	Écart type	4,14	6,42	0,67
	Nombre de publications	59	137	35

* Publications dans le domaine de la sécurité.

** Publications retenues par cette revue systématique.

Parmi les dix pays qui comptent le plus important nombre d'auteures et d'auteurs, sept sont des pays occidentaux et près de la moitié du total des auteures et auteurs provient des États-Unis (23,7 %) et du Royaume-Uni (22,6 %) (Tableau 9). L'importance d'auteures et d'auteurs occidentaux est cohérente avec le nombre d'évaluations identifiées dans notre revue systématique. Les auteures et auteurs non occidentaux proviennent de l'Indonésie, du Pakistan et du Kenya. Les deux premiers correspondent

respectivement à 42,6 % et 25 % des auteures et auteurs de l'Asie. Les auteures et auteurs kenyans, pour leur part, correspondent à 52,6 % des auteures et auteurs africains et sont suivis par les Marocains (15,8 %) et les Nigériens (10,5 %).

Tableau 9. Les dix pays ayant le nombre le plus important d'auteurs et d'auteurs

Pays	n	%
États-Unis	92	23,7 %
Royaume-Uni	88	22,6 %
Indonésie	29	7,5 %
Australie	27	6,9 %
Pakistan	17	4,4 %
Canada	16	4,1 %
Pays-Bas	16	4,1 %
Kenya	10	2,6 %
Belgique	7	1,8 %
Italie	6	1,5 %

À la lumière de cette analyse, il est évident que l'évaluation de programmes de prévention de l'extrémisme violent est un territoire largement colonisé par les pays occidentaux. Le cas de l'Afrique est le plus frappant. Plusieurs programmes africains ont été financés par l'Agence de développement international des États-Unis (USAID), mis en place vraisemblablement par des organismes américains et évalués par des évaluatrices, évaluateurs, chercheuses et chercheurs américains qui, dans plusieurs cas, comme nous le verrons dans la section des limites, ne parlaient pas la langue du pays en question et dont les publications issues de ces évaluations ont été rédigées, dans de nombreux cas, dans une langue qui n'était pas la langue du pays où le programme a été évalué.

3.2.3 Disciplines représentées dans le champ de l'évaluation

Moins du quart des auteures et auteurs (21,6 %) (Tableau 10) dans le domaine de l'évaluation ont pour discipline principale les sciences politiques, ce qui est vrai pour l'ensemble des régions et assez conforme avec les tendances observées dans le champ plus général des études sur le terrorisme et, dans une moindre mesure, celui des études sur l'extrémisme violent (Schuurman, 2018). Assez loin derrière, les autres disciplines les plus représentées dans le sous-champ de l'évaluation sont la santé mentale (11,1 %), la sociologie et le travail social (9 %), l'éducation (8 %), les études sur la paix et la sécurité (7,7 %), la criminologie (6,7 %) et, dans une moindre mesure, la santé publique/médecine (4,1 %).

Certaines disciplines qui peuvent sembler d'une pertinence particulière pour l'évaluation de programmes en prévention sont très peu ou sous-représentées. C'est notamment le cas de la communication (2,6 %) qui apparaît pourtant utile pour mesurer l'impact des campagnes de sensibilisation primaire et secondaire, hors ligne et en ligne. C'est aussi le cas de la théologie (0,8 %) qui est souvent évoquée dans les programmes de prévention secondaire et tertiaire, en particulier lorsqu'on considère l'extrémisme violent associé au djihadisme.

Tableau 10. Disciplines des auteures et auteurs des publications retenues

Disciplines	n	%
Sciences politiques	84	21,6 %
Santé mentale (Psychologie/ Psychiatrie)	43	11,1 %
Sociologie/ Travail social	35	9,0 %
Éducation	31	8,0 %
Sécurité / Études sur la paix	30	7,7 %
Criminologie / Police	26	6,7 %
Santé publique/ Médecine	16	4,1 %
Économie	13	3,3 %
Communication / Littérature	10	2,6 %
Sciences juridiques	8	2,1 %
Anthropologie	6	1,5 %
Administration / Management	6	1,5 %
Démographie / Géographie	3	0,8 %
Histoire	3	0,8 %
Études théologiques	3	0,8 %
Philosophie	3	0,8 %

Si ces résultats s'appliquent de manière générale à l'ensemble des régions, on note toutefois quelques disparités régionales pour certaines disciplines qui méritent d'être soulignées (Tableau 11). Si les sciences politiques sont omniprésentes dans l'ensemble des régions, ce n'est pas le cas de la sociologie et du travail social. Surtout présente en Afrique (25 %) et en Europe (16,8 %), cette dernière l'est beaucoup moins en Amérique du Nord (6,6 %), en Asie (2 %) et en Océanie (3,8 %). À l'exception de l'Amérique du Nord (11 %), très peu d'auteurs et d'auteurs sont issus de la santé publique/médecine en Europe (2,9 %), en Asie (4,1 %), en Afrique (0 %) et en Océanie (0 %). De la même manière, aucun ou très peu d'auteurs et d'auteurs proviennent du domaine de la santé mentale en Afrique (0 %) et en Océanie (3,8 %) alors que cette discipline est davantage présente en Amérique du Nord (16,5 %), en Europe (14,6 %) et en Asie (14,3 %). Les chiffres combinés de la santé, médecine et psychologie peuvent toutefois sembler, dans l'ensemble, assez bas au regard de l'importance de ces disciplines en matière de prévention secondaire et tertiaire.

Si elle est très présente en Océanie (26,9 %), l'éducation l'est moins en Europe (10,2 %) et en Asie (12,2 %) et presque absente en Amérique du Nord (4,4 %) et en Afrique (0 %). Cela peut surprendre au regard de l'importance des programmes dans le secteur éducatif en matière de prévention primaire et secondaire et laisse penser que les évaluations y sont donc effectuées par des personnes d'autres champs disciplinaires.

Alors qu'elle représente près du quart des auteures et auteurs (26,9 %) en Océanie, 11 % en Amérique du Nord et 6,6 % en Europe, la criminologie est absente en Afrique (0 %) et en Asie (0 %).

Tableau 11. Nombre d'auteures et d'auteurs par discipline et par continent

Discipline		Continent				
		Afrique	Amérique du Nord	Asie	Europe	Océanie
Anthropologie	n	0	1	0	4	1
	%	0,0 %	1,1 %	0,0 %	2,9 %	3,8 %
Criminologie / Police	n	0	10	0	9	7
	%	0,0 %	11,0 %	0,0 %	6,6 %	26,9 %
Démographie/ Géographie	n	0	1	0	1	1
	%	0,0 %	1,1 %	0,0 %	0,7 %	3,8 %
Économie	n	2	0	9	2	0
	%	12,5 %	0,0 %	18,4 %	1,5 %	0,0 %
Éducation	n	0	4	6	14	7
	%	0,0 %	4,4 %	12,2 %	10,2 %	26,9 %
Histoire	n	0	0	1	2	0
	%	0,0 %	0,0 %	2,0 %	1,5 %	0,0 %
Santé mentale (Psychologie/ Psychiatrie)	n	0	15	7	20	1
	%	0,0 %	16,5 %	14,3 %	14,6 %	3,8 %
Santé publique/ Médecine	n	0	10	2	4	0
	%	0,0 %	11,0 %	4,1 %	2,9 %	0,0 %
Sciences juridiques	n	1	1	1	5	0
	%	6,3 %	1,1 %	2,0 %	3,6 %	0,0 %
Sciences politiques	n	4	25	12	39	4
	%	25,0 %	27,5 %	24,5 %	28,5 %	15,4 %
Sociologie / Travail social	n	4	6	1	23	1
	%	25,0 %	6,6 %	2,0 %	16,8 %	3,8 %
Sécurité / Études sur la paix	n	3	11	5	7	3
	%	18,8 %	12,1 %	10,2 %	5,1 %	11,5 %
Études théologiques	n	0	1	0	2	0
	%	0,0 %	1,1 %	0,0 %	1,5 %	0,0 %
Communication / Littérature	n	1	1	5	3	0
	%	6,3 %	1,1 %	10,2 %	2,2 %	0,0 %
Philosophie	n	0	1	0	1	1
	%	0,0 %	1,1 %	0,0 %	0,7 %	3,8 %
Administration / Management	n	1	4	0	1	0
	%	6,3 %	4,4 %	0,0 %	0,7 %	0,0 %

3.2.4 Omniprésence du milieu de la recherche dans les professions des auteures et auteurs

Un peu plus de la moitié des auteures et auteurs (63,3 %) occupent un emploi de professeure, professeur ou de chercheuse, chercheur dans le milieu universitaire (Tableaux 12 et 13). S'il est moins élevé en Afrique (30,8 %), ce chiffre grimpe jusqu'à 71,2 % en Europe et 84 % en Océanie. Les chercheuses et chercheurs d'établissement des secteurs tiers (fondation, ONG, *think tank*) représentent, quant à eux, 16,1 % de l'ensemble des auteures et auteurs. Les personnes chercheuses du milieu gouvernemental comptent enfin pour 2,5 % des auteures et auteurs. En d'autres termes, le milieu de la recherche comprend plus des deux tiers des auteures et auteurs en matière d'évaluation.

Bien qu'en nombre très inférieur, les personnes consultantes sont également présentes, qu'elles soient indépendantes (5,7 %) ou dans une firme privée (5,4 %) (Tableaux 12 et 13). Toutefois, on les retrouve surtout en Asie (31,9 %), en Afrique (30,8 %) et, dans une moindre mesure, en Océanie (9,4 %), mais très peu en Amérique du Nord (5,6 %) et en Europe (4,3 %). Du côté des milieux de pratique, les directrices et directeurs/coordonnatrices et coordonnateurs et les managers comptent pour 6,6 % des auteures et auteurs. Ils sont toutefois beaucoup plus nombreux en Amérique du Nord (14,6 %) qu'en Europe (4,3 %), en Afrique (7,7 %), en Asie (0 %) ou en Océanie (0 %). Les personnes intervenantes sont, quant à elles, tout simplement absentes (moins de 0,3 %).

Tableau 12. Nombre d'auteures et d'auteurs par profession

Professions	n	%
Chercheuse, chercheur universitaire (Professeure, professeur, personnes doctorantes et postdoctorantes)	200	63,3 %
Chercheuse, chercheur d'établissement - secteur tiers (fondation, ONG, think tank, etc.)	51	16,1 %
Directrice, directeur / coordonnatrice, coordonnateur / manager	21	6,6 %
Consultante, consultant (indépendant)	18	5,7 %
Consultante, consultante (firme privée)	17	5,4 %
Chercheuse, chercheur dans un organisme gouvernemental	8	2,5 %
Intervenante, intervenant	1	0,3 %

Tableau 13. Nombre d'auteures et d'auteurs par profession et par continent

		Continent				
		Afrique	Amérique du Nord	Asie	Europe	Océanie
Chercheuse, chercheur universitaire	n	4	50	26	99	21
	%	30,8 %	56,2 %	55,3 %	71,2 %	84,0 %
Chercheuse, chercheur d'établissement - secteur tiers	n	4	18	6	23	0
	%	30,8 %	20,2 %	12,8 %	16,5 %	0,0 %
Directrice, directeur coordonnatrice, coordonnateur manager	n	1	13	0	6	0
	%	7,7 %	14,6 %	0,0 %	4,3 %	0,0 %
Consultante, consultant (indépendant)	n	3	3	4	6	1
	%	23,1 %	3,4 %	8,5 %	4,3 %	4,0 %
Consultante, consultant (boite privée)	n	1	2	11	0	17
	%	7,7 %	2,2 %	23,4 %	0,0 %	5,4 %
Chercheuse, chercheur dans un organisme gouvernemental	n	0	3	0	4	1
	%	0,0 %	3,4 %	0,0 %	2,9 %	4,0 %
Intervenante, intervenant	n	0	0	0	1	0
	%	0,0 %	0,0 %	0,0 %	0,7 %	0,0 %

L'un des éléments importants d'un processus d'évaluation est sans aucun doute la mise sur pied de l'équipe d'évaluation. Sa diversité et sa représentativité au regard du programme et des actrices et acteurs sont des éléments importants. Or, l'un des constats les plus importants de cette revue concernant les auteures et auteurs des évaluations réside dans la très forte représentation du milieu de la recherche en matière d'évaluation de l'extrémisme violent. Dans la section concernant le type de publication, nous avons remarqué l'importance de la littérature grise dans l'évaluation de ce type de programme. Nous avons évoqué l'idée selon laquelle la présence de ce type de littérature pouvait envisager un développement au-delà de l'espace académique, cependant ces résultats démontrent, au contraire, que l'académie est bel et bien présente même si elle ne vise pas spécifiquement la publication dans des revues spécialisées. Les chercheuses et chercheurs ne sont toutefois pas, en principe, celles et ceux qui, majoritairement, mettent en œuvre les programmes de prévention. Cela soulève plusieurs questions importantes. Par qui et pour qui sont réalisées ces évaluations ?

Sont-elles destinées d'abord aux milieux de la recherche eux-mêmes et, dans une certaine mesure, aux bailleurs de fonds ? Dans quelle mesure les milieux de pratique, qui sont largement absents parmi les auteures et auteurs, peuvent-ils se sentir interpellés par des évaluations réalisées par des chercheuses et chercheurs donc, de prime abord, plus éloignées de leurs préoccupations ?

Dans quelle mesure ces évaluations sont-elles adaptées, en termes d'objectifs et de langage, à l'usage que pourraient en faire des personnes intervenantes ? Ces questions sont encore plus importantes si l'on considère, comme nous l'avons indiqué dans la partie précédente, qu'une partie des évaluations ont été rédigées en anglais pour des programmes mis en place dans des pays non anglophones. En plus, comme nous le verrons dans la partie concernant les limites, l'une des difficultés rencontrées par les chercheuses et chercheurs sur le terrain a été le problème de traduction pour des chercheuses et chercheurs qui ne parlaient pas la langue du pays où ils faisaient l'évaluation.

3.3 COMMENT CES PROGRAMMES ONT-ILS ÉTÉ ÉVALUÉS ? UNE QUESTION DE DESIGN MÉTHODOLOGIQUE

Cette section porte sur les dimensions méthodologiques de la procédure d'évaluation des programmes évalués. Ces dimensions méthodologiques comprennent le type d'évaluation ciblée par les études, le type d'évaluateur, l'approche globale, le type de manipulation des variables utilisées, le temps des observations, les outils de collecte de données, les échantillons et les types d'indicateurs utilisés.

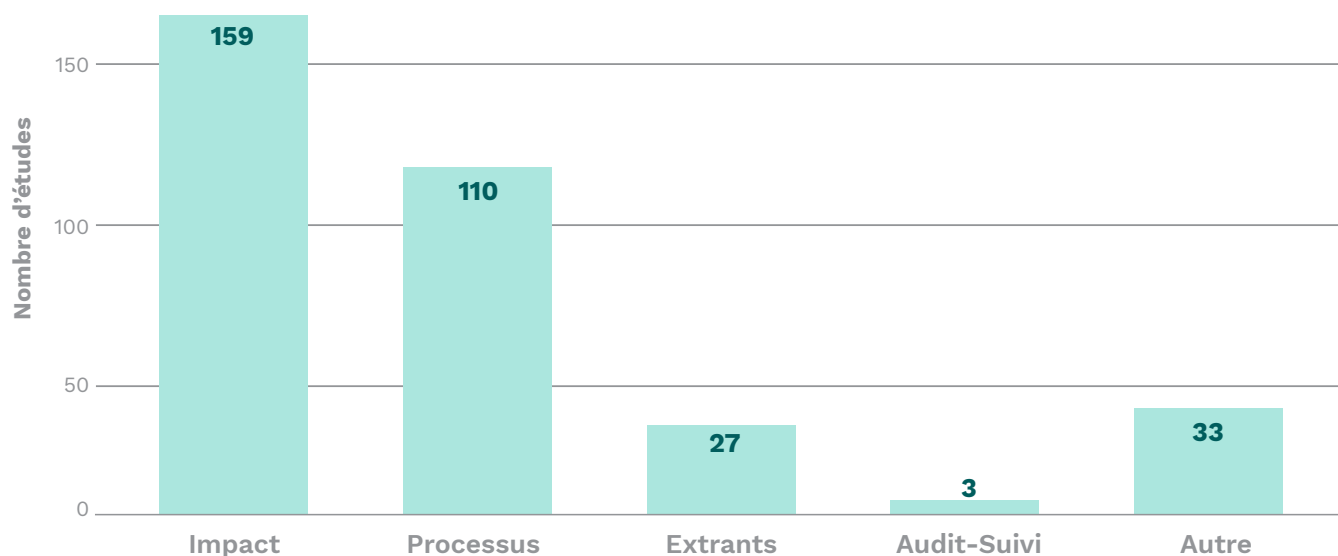
3.3.1 Quel type d'évaluation ces études visaient-elles ?

Nous avons défini cinq types d'évaluation selon leurs objectifs³³ : d'impact, de processus, d'extrants, audit et suivi. Comme la Figure 11 le montre, la plupart des études ont visé l'évaluation des effets que l'intervention a eus sur la population cible (n = 159) et l'évaluation de la mise en place des programmes (processus, n = 110), un résultat qui suit la même ligne que celle des recensions précédentes (Bellasio et al., 2018 ; Feddes, 2015 ; Mastroe et Szmania, 2016). Une grande partie de ces études (59,4 %) ciblait exclusivement un de ces cinq types, suivies par celles qui en avaient deux (32,4 %). Les études qui ont ciblé

un seul type d'évaluation se sont concentrées en grand nombre sur les évaluations d'impact (52,2 %) ou sur les évaluations de processus (35,5 %). Dans le cas des études qui ont ciblé deux types d'évaluation, 8 études sur 10 ont ciblé l'impact et le processus. Les évaluations d'extrants, les suivis et les audits ont ainsi été majoritairement des composantes d'une évaluation plus globale qui ciblait un ou deux de ces types principaux (impact et processus). Malgré cette conclusion, les évaluations d'extrants, c'est-à-dire la description de ce qui a été réalisé par rapport à ce qui était prévu, semblent être une composante importante, mais non fondamentale de l'évaluation des programmes dans ce domaine.

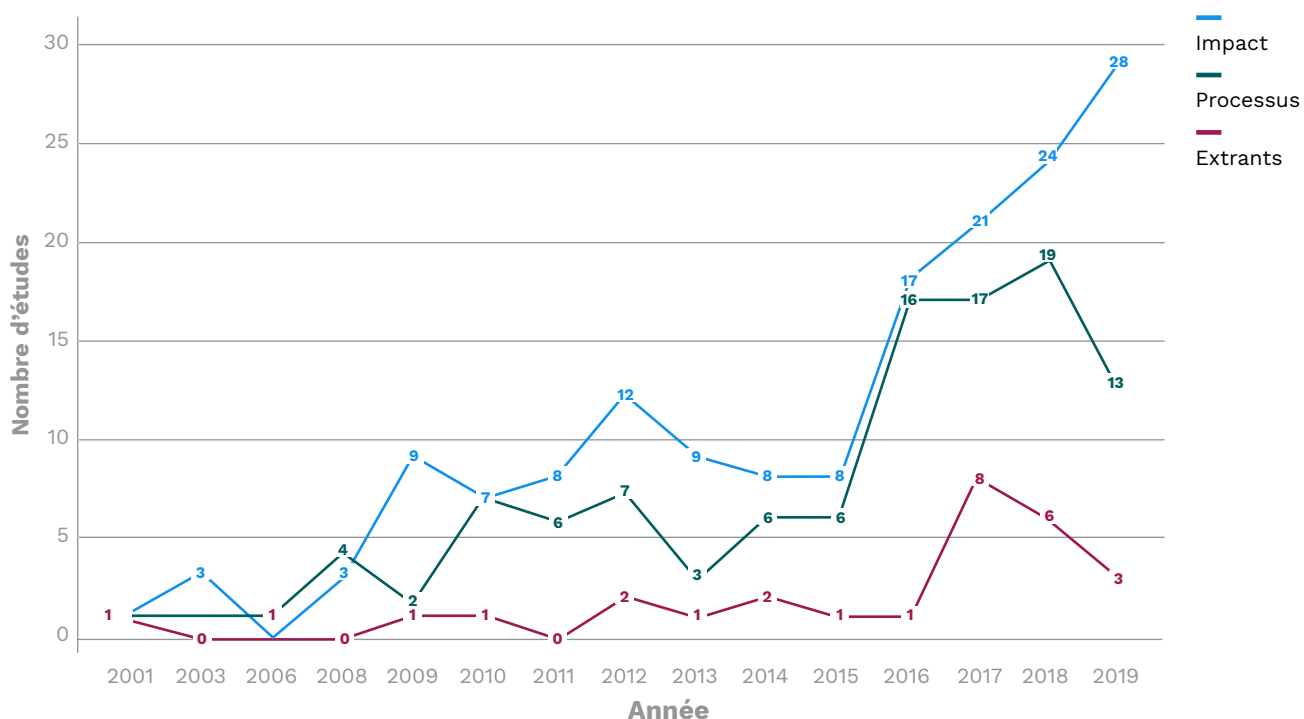
³³ Une définition de cette typologie se trouve dans les annexes.

Figure 11. Types d'évaluation



À la lumière des objectifs de l'évaluation, nous pouvons observer que le point tournant en 2016 ne concerne principalement que les études visant à évaluer les effets de l'intervention (Figure 11). Les évaluations de processus, ayant connu une augmentation importante entre 2015 et 2016, demeurent relativement stables à partir de cette dernière année et diminuent dans la dernière année. Les évaluations d'extrants, pour leur part, connaissent une augmentation temporaire en 2017 et, ensuite, une diminution progressive. Il faut préciser que nous avons codé ces types d'évaluation lorsqu'ils ont été évoqués directement par les auteures et auteurs. Comme nous le verrons plus tard, ces éléments de déclaration ne sont pas forcément associés à un design méthodologique cohérent afin d'atteindre ces objectifs. De cette façon, cette augmentation progressive des évaluations d'impact à partir de 2016 peut davantage faire preuve d'une augmentation de l'intention à évaluer les effets de l'intervention qu'une évaluation effective de ces effets.

Figure 12. Types d'évaluation par année



Le Tableau 14 présente davantage d'informations concernant les trois principaux objectifs de l'évaluation par continent, par niveau de prévention et par l'extrémisme ciblé par les programmes. Par continent, la situation est assez diversifiée. Bien que les évaluations d'impact soient toujours plus présentes dans tous les continents, la proportion varie d'une région à une autre. L'Océanie (85,7 %) et l'Afrique (82 %) sont les deux continents qui détiennent la plus grande proportion d'évaluations dites d'« impacts », comparativement aux autres continents. Cependant, l'équilibre impact/processus est très différent dans les deux cas. Dans le cas de l'Afrique, par exemple, pour chaque évaluation de processus, 2,2 évaluations d'impact ont été réalisées, tandis qu'en Océanie, ce rapport est de 1,5 qui est proche du ratio global entre les études d'impact et les études de processus (1,4). Ce ratio nous montre ainsi le profil de l'évaluation par continent. Dans le reste des continents, cette relation est plus équilibrée (Asie=1,3; Europe=1,2; Amérique du Nord=1,7), l'Europe étant en situation d'une relative

parité. La disproportion entre les deux types d'objectifs dans le cas du continent africain peut être le résultat du type d'évaluateur et de l'influence du bailleur de fonds. Comme nous l'avons vu dans le cas de l'étude qualitative internationale que nous avons menée en parallèle à cette revue systématique (Madriaza et al., 2021), les bailleurs de fonds semblent être moins intéressés par la mise en place de ces programmes que pour les effets qu'ils produisent. Les intervenantes et intervenants, en revanche, montrent plus d'intérêt sur les évaluations qualitatives du processus que pour les résultats des méthodes quantitatives d'évaluation d'impact. Dans le cas de l'Afrique, comme nous l'avons vu dans les deux premières sections de cette revue systématique, les bailleurs de fonds proviennent souvent de l'extérieur de l'Afrique et la plupart des personnes évaluatrices proviennent également des pays occidentaux. Il en résulte que l'évaluation est souvent planifiée depuis l'extérieur du programme, à partir d'une perspective « top-down », où les actrices et acteurs du terrain ont très peu à dire.

Tableau 14. Type d'objectif de l'évaluation par continent, par niveau de prévention et par type d'extrémisme

		Impact		Processus		D'extrants	
		n	%	n	%	n	%
Total		159	72,6 %	110	50,2 %	27	12,3 %
Continent	Afrique	41	82,0 %	19	38,0 %	11	22,0 %
	Amérique du Nord	10	55,6 %	6	33,3 %	3	16,7 %
	Asie	32	76,2 %	24	57,1 %	3	7,1 %
	Europe	67	68,4 %	57	58,2 %	9	9,2 %
	N/A	3	75,0 %	0	0,0 %	1	25,0 %
	Océanie	6	85,7 %	4	57,1 %	0	0,0 %
Niveau de prévention	Primaire	36	75,0 %	16	33,3 %	10	20,8 %
	Primaire ciblée	71	79,8 %	47	52,8 %	8	9,0 %
	Secondaire	42	68,9 %	33	54,1 %	13	21,3 %
	Tertiaire	29	63,0 %	33	71,7 %	7	15,2 %
	Générale	16	66,7 %	10	41,7 %	3	12,5 %
Type d'extrémisme	Extrême droite	16	80,0 %	6	30,0 %	2	10,0 %
	Islamisme	61	72,6 %	47	56,0 %	10	11,9 %
	Non spécifique	91	71,7 %	62	48,8 %	16	12,6 %

Ce ratio change également en fonction du niveau de prévention et du type d'extrémisme que le programme cible. Dans le premier cas, ce sont les programmes plus universels, ceux qui misent davantage sur l'évaluation d'impact que sur l'évaluation du processus (prévention primaire = 2,3; prévention générale = 1,6; prévention primaire ciblée=1,5). Les études qui évaluent des

programmes plus spécifiques, en revanche, maintiennent un équilibre plus prononcé (prévention secondaire=1,3) ou la tendance est inversée (prévention tertiaire = 0,9). Dans l'introduction de ce rapport, nous avons déjà évoqué les problèmes pour mesurer les effets dans le cas de programmes de prévention tertiaire et les enjeux éthiques pour l'expérimentation dans ce domaine. Comme nous

le verrons dans les sections consacrées aux designs et à l'évaluation de la qualité méthodologique, les designs quantitatifs plus sophistiqués (expérimentaux et quasi expérimentaux) sont plus présents dans les niveaux plus universels, car ces problèmes méthodologiques et éthiques sont moins importants. Quant au type d'extrémisme ciblé par les programmes évalués, on observe que la tendance est relativement différente par rapport à la spécificité du programme. Les programmes qui ciblent l'extrémisme islamiste (ratio = 1,3) et qui ne ciblent pas spécifiquement un type d'extrémisme en particulier (ratio = 1,5) gardent un certain équilibre entre les deux types principaux d'évaluation par rapport aux programmes qui ciblent l'extrême droite (ratio = 2,7). L'équilibre relatif entre les

deux principaux types d'évaluations de programmes qui ciblent l'islamisme radical est compréhensible étant donné les enjeux éthiques des évaluations quantitatives et en raison de l'intérêt possible des évaluatrices et évaluateurs pour connaître la réalité à partir de la voix des actrices et acteurs du terrain, qui est souvent le cas des évaluations de processus, notamment dans un contexte où la communauté musulmane a été fortement stigmatisée. La disproportion dans le cas des évaluations de programmes d'extrême droite est surprenante. Serait-ce le résultat de l'insistance de la littérature sur l'absence d'évaluations de ce type de programme ? C'est difficile à dire.

3.3.2 Des évaluatrices et évaluateurs indépendants ?

Si, dans la section 3.2 de ces résultats, nous avons brossé un portrait individuel des auteures et auteurs des études, ici, nous nous concentrons sur l'indépendance de la personne évaluatrice ou du groupe d'évaluation par rapport au programme évalué. Nous avons ainsi défini deux grandes catégories : l'évaluation externe et l'évaluation interne (Tableau 15). La première concerne les évaluations conduites par des services ou des personnes extérieures au bailleur de fonds et à l'organisation responsable de la mise en œuvre du programme. Les évaluations internes sont, en revanche, dans ce contexte, les évaluations réalisées par celles et ceux qui ont la responsabilité de concevoir et de mettre en œuvre une action ou un programme ou dans laquelle ou lequel

participent également les organismes bailleurs ou leurs partenaires. **Les trois quarts des équipes d'évaluation ont été composés par des évaluatrices et évaluateurs externes à l'organisation** (n=158) et, seulement, 51 études ont été réalisées par des évaluations internes. Dans le reste des études, nous n'avons pas pu déterminer la nature de l'équipe de personnes évaluatrices en raison du manque d'informations concernant les auteures et auteurs des études. Nous n'avons pas trouvé non plus d'études de nature participative, c'est-à-dire où toutes les parties prenantes ou une grande partie d'entre elles (y compris les bénéficiaires, les personnes intervenantes et chercheuses) ont collaboré pour mener l'évaluation.

Tableau 15. Type d'évaluatrice, d'évaluateur par continent, par niveau de prévention et par type d'extrémisme

		Externe		Interne	
		n	%	n	%
Total		158	72,10 %	51	23,30 %
Continent	Afrique	35	76,10 %	11	23,90 %
	Amérique du Nord	11	64,70 %	6	35,30 %
	Asie	31	77,50 %	9	22,50 %
	Europe	74	77,90 %	21	22,10 %
	N/A	1	25,00 %	3	75,00 %
	Océanie	6	85,70 %	1	14,30 %
Niveau de prévention	Primaire	38	82,60 %	8	17,40 %
	Primaire ciblée	61	69,30 %	27	30,70 %
	Secondaire	42	72,40 %	16	27,60 %
	Tertiaire	31	73,80 %	11	26,20 %
	Générale	20	87,00 %	3	13,00 %
Type d'extrémisme	Extrême droite	11	57,90 %	8	42,10 %
	Islamisme	58	71,60 %	23	28,40 %
	Non spécifique	97	80,20 %	24	19,80 %

Le ratio entre les évaluations internes et les évaluations externes peut également être utile ici pour tracer ce portrait. Dans l'ensemble, pour chaque évaluation interne, 3,1 évaluations externes ont été réalisées. Considérant ce ratio global, la plupart des continents présentent un ratio légèrement au-dessus de ce chiffre (Afrique = 3,2 ; Asie = 3,4 ; Europe = 3,5), l'Océanie montre une préférence plus importante pour les évaluations externes (ratio = 6) tandis qu'en Amérique du Nord (ratio = 1,8), ce ratio est beaucoup moins important. La proportion plus faible dans le cas de l'Amérique du Nord peut être due à la nature plus expérimentale des programmes appliqués qui ont été évalués par les mêmes chercheuses et chercheurs qui les ont conçus.

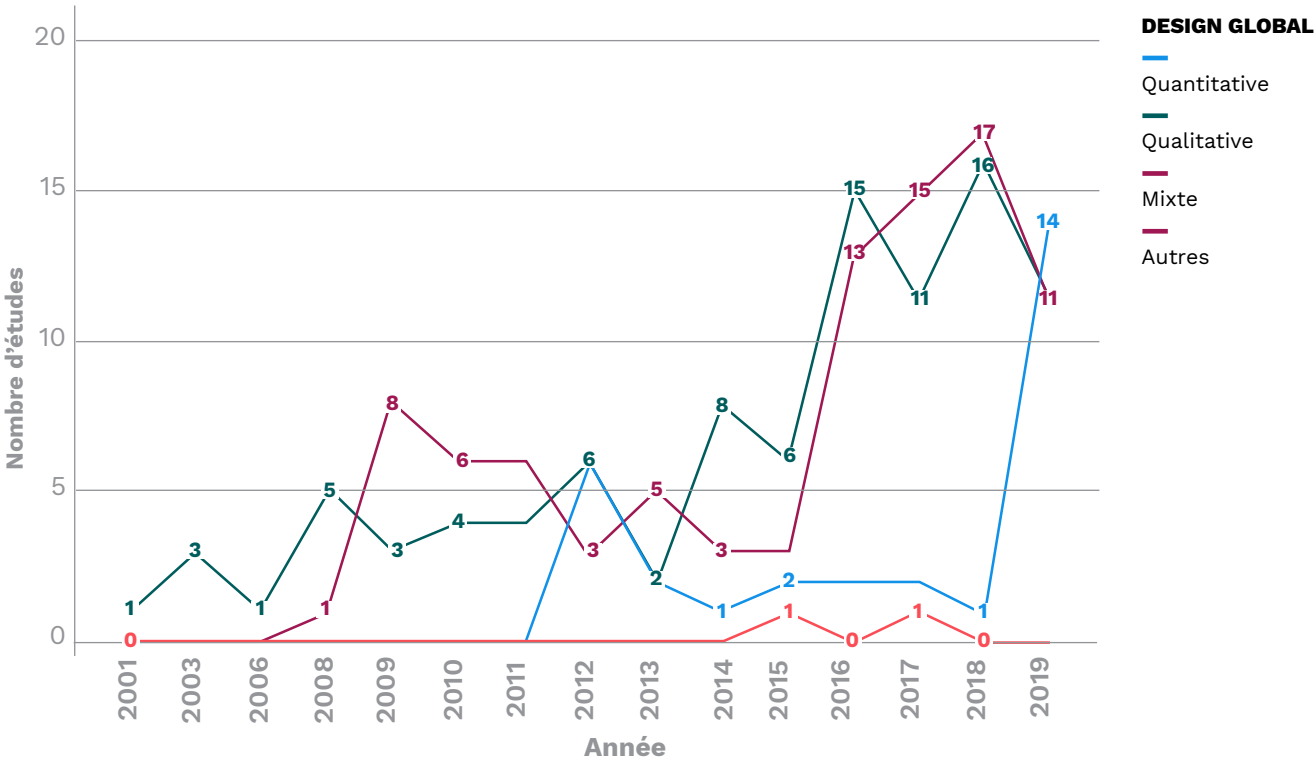
Par niveau de prévention, la proportion entre les évaluations internes et externes montre une tendance claire : les études qui évaluent des programmes plus universels ont un ratio plus élevé (prévention primaire = 4,8 ; prévention générale = 6,7) et celles qui ont ciblé des programmes plus spécifiques sont en dessous du ratio global (prévention primaire ciblée = 2,3 ; prévention secondaire = 2,6 ; prévention tertiaire = 2,8). Cette différence peut s'expliquer par la nature plus sensible de l'évaluation des programmes plus spécifiques qui nécessitent des points de vue divers autant internes qu'externes pour bien mener l'évaluation. Cette situation se confirme si nous observons les évaluations par type d'extrémisme ciblé par le programme évalué : les évaluations des programmes plus spécifiques (extrême

droite = 1,4 ; extrémisme islamiste = 2,5) ont un ratio en dessous du ratio global tandis que les programmes non spécifiques montrent une proportion plus importante d'évaluations externes par rapport aux évaluations internes (ratio = 4,2).

3.3.3 Quelle approche globale a été utilisée ?

L'approche globale des études fait référence à l'utilisation des méthodologies qualitatives, quantitatives ou mixtes (quantitatives et qualitatives) (Figure 13). Les approches purement qualitatives (n = 96 ; 43,8 %) et mixtes (n = 91 ; 41,6 %) sont les designs les plus utilisés dans l'évaluation de ce type de programmes. Seules 30 études ont été exclusivement quantitatives (13,7 %). Les évaluations quantitatives ont cependant connu une augmentation exponentielle dans la dernière année de l'étude, en 2019, où elles dépassent pour la première fois les évaluations mixtes et qualitatives. Cette croissance est attribuable à l'augmentation des publications académiques, mais elle peut également s'expliquer par l'intérêt grandissant des dernières années quant à l'évaluation d'impact (voir la section précédente) qui est habituellement associée à une approche quantitative. Proportionnellement, l'Océanie (28,6 %) est la région qui a le plus utilisé ce type d'évaluation quantitative (Tableau 16), suivi de l'Amérique du Nord (16,7 %).

Figure 13. Évolution des approches méthodologiques utilisées par année



La progression des études exclusivement qualitatives a commencé en 2014, légèrement en avance sur le reste des études. Toutefois, ces dernières années, cette progression a été dépassée par les études mixtes et, en particulier, par les études quantitatives réalisées dans l'année 2019. Cette situation peut être associée à la stagnation des évaluations de processus des dernières années qui utilisent principalement ce type d'approche. Les approches qualitatives sont probablement l'une des caractéristiques des évaluations en Europe qui représentent 53,1% du total des études menées dans le vieux continent.

Un autre résultat intéressant est l'importance des approches mixtes qui constitue, comme nous l'avons vu dans la première section du rapport, une recommandation récurrente des expertes et experts. La progression des

designs mixtes connaît une augmentation considérable à partir de l'année 2016. Ils sont proportionnellement plus présents en Océanie (71,4 %) et en Afrique (50 %) et moins en Europe (34,7 %) (Tableau 16). Les équipes de Bellasio (2018) et de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021) avaient déjà observé une augmentation des études mixtes dans les dernières années. Cependant, les résultats de notre revue systématique, quant à cette dernière approche, révèlent une importance encore plus grande alors qu'ils couvrent une période semblable à celle des études précitées. Cela s'explique sans doute en partie par le biais de publication des recensions précédentes qui se sont concentrées en majorité sur la littérature académique. Ainsi, 70 des 91 études mixtes de notre revue systématique ont été repérées dans la littérature grise.

Tableau 16. Approche globale par continent et par type d'évaluation

		Quantitative		Qualitative		Mixte	
		n	%	n	%	n	%
Total		30	13,7 %	96	43,8 %	91	41,6 %
Continent	Afrique	6	12,0 %	18	36,0 %	25	50,0 %
	Amérique du Nord	3	16,7 %	6	33,3 %	8	44,4 %
	Asie	5	11,9 %	20	47,6 %	17	40,5 %
	Europe	12	12,2 %	52	53,1 %	34	34,7 %
	N/A	2	50,0 %	0	0,0 %	2	50,0 %
	Océanie	2	28,6 %	0	0,0 %	5	71,4 %
Type d'évaluation	Impact	29	18,5 %	49	30,8 %	81	50,9 %
	Processus	1	0,9 %	66	60,0 %	43	39,1 %
	D'Extrants	2	7,4 %	9	33,3 %	14	51,9 %

Un dernier élément à considérer est la relation entre le type d'approche méthodologique globale et le type d'évaluation réalisée (Tableau 16). Nous avons noté plus haut que les objectifs de l'évaluation ne correspondaient pas toujours aux résultats obtenus ou aux outils utilisés. Bien qu'il soit possible d'évaluer l'impact d'un programme sur la base d'informations analysées de manière qualitative, l'évaluation des effets d'un programme est souvent réalisée à l'aide de méthodes quantitatives. L'idée inverse est également possible, c'est-à-dire évaluer quantitativement le processus de mise en œuvre d'un programme, bien que méthodologiquement ce processus soit souvent évalué à l'aide de techniques qualitatives, car il implique l'exploration d'éléments qui ne sont pas toujours connus à l'avance par les évaluatrices et évaluateurs. Dans le cas des études retenues dans cette recherche, cette logique demeure relativement inaltérée, notamment en ce qui concerne l'évaluation de processus

qui se fait presque exclusivement par des méthodologies qualitatives ou mixtes (99,1%)³⁴. Dans le cas de l'évaluation d'impact, cette affirmation est plus nuancée : 30,8 % des études qui affirment vouloir évaluer l'impact ont utilisé exclusivement des techniques qualitatives. Bien qu'un nombre important de ces études aient documenté des résultats cohérents avec l'objectif d'évaluer les effets du programme, dans le reste des études, nous n'avons pas pu vérifier cette cohérence. En d'autres termes, les objectifs initiaux indiqués dans la méthodologie ou dans l'introduction de l'étude, comme nous le verrons dans l'évaluation de la qualité des méthodes utilisées, ne sont pas toujours atteints avec les résultats obtenus. Dans ces cas, l'évaluation d'impact semble être un concept qui a été relativement vidé de sa définition originale et qui est donc associé de façon générale à tout type d'évaluation réalisée.

³⁴ Une seule étude a indiqué vouloir évaluer le processus alors qu'elle n'a utilisé que des outils quantitatifs (Broadbent, 2013). L'auteur indique dans sa méthodologie des questions qui pointent vers une évaluation de processus, mais ne rapporte que des données sur les effets et sur la satisfaction du programme.

3.3.4 Combien d'évaluations ont utilisé des méthodes expérimentales ?

Dans cette section, nous nous intéressons à la mesure dans laquelle les études d'évaluation incluses dans cette revue systématique ont été conçues comme des expériences scientifiques, c'est-à-dire si ces études étaient expérimentales ou quasi expérimentales³⁵ (Encadré 3). Sans surprise, la majorité des études évaluatives (n=157) ont utilisé des designs autres que ces derniers, étant pour la plupart des études qualitatives ou quantitatives descriptives : une tendance qui est semblable à celles constatées dans d'autres recensions. Cependant, à la différence de ces dernières, cette revue systématique a identifié 54 études quasi expérimentales et 6 études expérimentales.

54 études quasi expérimentales et 6 études expérimentales ont été identifiées dans cette revue systématique.

Normalement, ce type d'étude est utilisé pour évaluer l'impact des programmes sur les personnes participantes et c'est, en effet, l'objectif de toutes

les études expérimentales et quasi expérimentales identifiées dans cette revue. Une étude expérimentale et douze études quasi expérimentales ont par ailleurs évalué le processus de mise en place des programmes par l'intermédiaire d'analyses qualitatives (études mixtes). Le nombre d'études expérimentales et quasi expérimentales identifiées est encourageant pour le domaine, notamment en ce qui concerne l'évaluation des effets de ces programmes. L'étude de Bellasio et ses collègues (2018) n'avait, par exemple, identifié que six études ayant recouru à des mesures *ex post* alors que celle des équipes de Hassan (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021), plus récente, en a identifié seulement quatre³⁶. Hormis la recherche approfondie que nous avons faite sur le plan de la littérature grise, ce nombre relativement bas peut s'expliquer, dans le cas de leurs deux dernières études, par des critères d'inclusion plus restreints. Comme nous le verrons dans la section consacrée à la qualité méthodologique, le recours aux méthodes expérimentales ou quasi expérimentales n'est malheureusement pas une garantie de la qualité méthodologique de l'étude. Cet aspect a été négligé dans le cas de l'étude de Bellasio et ses collègues (2018).

Encadré 3. Deux exemples d'évaluations expérimentales

The Effectiveness of an Educational Program for Developing Tolerance Values and Resistance to Intellectual Extremism at Secondary Level in Jordan (Al-Maqosi et al., 2019)

La présente étude visait à établir un programme éducatif efficace destiné aux élèves du secondaire, en Jordanie. Ce programme mettait l'accent sur les droits des non-musulmans dans l'Islam afin d'encourager le développement de valeurs de tolérance à l'égard des diverses religions, tout en visant le développement d'une résistance intellectuelle face aux idéologies extrémistes. À cet effet, les jeunes élèves ont assisté à dix sessions éducatives, totalisant dix heures, afin de les éduquer sur divers droits de la personne fondamentaux (dignité, sécurité, liberté de religion, liberté d'expression, justice, etc.) L'étude a été menée sur un échantillon de 48 élèves de la 11e année de l'école islamique Al-Arqam, au cours de l'année scolaire 2017-2018. Suivant une méthodologie expérimentale, l'échantillon fut divisé en deux groupes distincts : 23 élèves dans le groupe de traitement, 25 élèves dans le groupe contrôle. Afin de mesurer l'efficacité du programme éducatif, une échelle mesurant les valeurs de tolérance religieuse ainsi qu'une échelle mesurant la résistance à l'extrémisme furent créées par les chercheurs. Les résultats ont montré une différence statistiquement significative entre la performance moyenne du groupe expérimental et celle du groupe témoin dans les mesures de la tolérance religieuse et de la résistance à l'extrémisme. En effet, alors que les scores moyens des élèves du groupe contrôle demeuraient relativement semblables, les scores moyens des élèves ayant suivi le programme éducatif augmentèrent significativement.

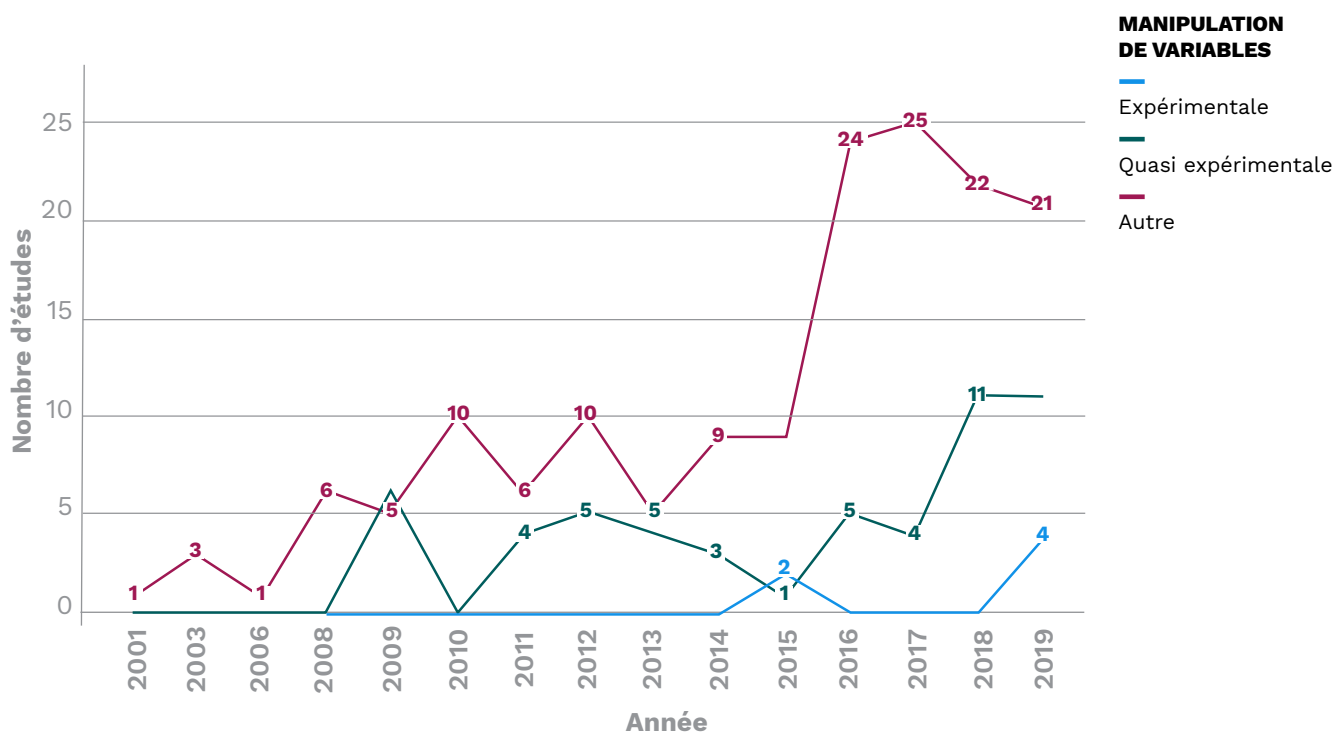
³⁵ Les études expérimentales nécessitent des mesures avant et après l'intervention et utilisent un groupe contrôle avec des participantes et participants sélectionnés au hasard. Les études quasi expérimentales, en revanche, ne répondent pas à au moins une de ces exigences (Salkind, 2010). Bien que certains auteurs et auteures incluent au sein des designs quasi expérimentaux les études qui ont exclusivement pris des mesures post-intervention et tenté de tester des hypothèses causales (Salkind, 2010), nous considérons ici, suivant Reichardt (2009), comme des études quasi expérimentales uniquement les études qui ont utilisé au minimum une mesure de comparaison, par exemple, un design pré-post avec un seul groupe ou un design post-test avec un groupe témoin non randomisé. Le reste des études quantitatives a été considéré dans cette recension comme étant des études quantitatives descriptives en raison des analyses descriptives non comparatives.

³⁶ L'étude de Carthy et ses collègues ont repéré des études expérimentales et quasi expérimentales. Toutefois, comme nous l'avons indiqué dans la première section de ce rapport, nous n'avons retenu aucune étude de cette revue parce qu'aucune étude ne ciblait directement la prévention de l'extrémisme violent.

Evaluation et Analytical Services (EAS). Project for The regional Peace and Governance Programs - Impact Evaluation of Peace Through Development II (P-DEV II) Radio Programming in Chad and Niger - FINAL REPORT (Finkel et al., 2015)³⁷

Les auteurs et auteure ont évalué la composante radio du programme « peace through development » cherchant à contrer l'extrémisme violent au Tchad, au Niger et au Burkina Faso. Cette étude s'est concentrée notamment sur les deux premiers pays. À cette fin, le programme concentre ses efforts afin de renforcer l'autonomie des jeunes, accroître les voix modérées et la résilience à l'extrémisme violent, ainsi que renforcer le gouvernement local. Pour ce faire, ce dernier misa sur le renforcement de la capacité des stations de radio locales à produire et à diffuser des contenus visant à contrer l'extrémisme violent en leur fournissant des équipements, une assistance technique et la formation de leur personnel. L'impact du programme radio fut évalué de manière longitudinale chez 750 jeunes Tchadiens âgés de 15 à 30 ans. Ainsi, une entrevue fut conduite auprès de chacun de ces jeunes avant l'implantation du programme, alors qu'une autre entrevue eut lieu plusieurs mois après. De plus, 525 des jeunes interviewés reçurent des messages d'encouragement et de rappel concernant l'écoute du programme radio, alors que les 225 autres n'en reçurent aucun. Les résultats ont indiqué que l'écoute de la radio augmentait l'insatisfaction à l'égard de la vie dans les deux pays. Au Tchad, l'écoute de la radio a également suscité un intérêt accru pour les affaires locales et la politique, en plus d'avoir des effets positifs quant à la diversité et l'inclusion et d'avoir contribué à réduire le soutien à la violence. Au Niger, contrairement aux résultats obtenus au Tchad, aucun effet significatif ne fut observé quant à la diversité et à l'inclusion ainsi qu'au soutien à la violence. Des effets significatifs différents furent obtenus, notamment la hausse de confiance envers le gouvernement local.

Figure 14. Évolution des designs expérimentaux par année



Les études expérimentales sont apparues, notamment, au cours de la dernière année considérée par cette revue systématique (voir Figure 14), ce qui peut expliquer son absence dans les autres recensions. Ce n'est pas le cas des études quasi expérimentales qui ont été relativement présentes depuis 2009 et qui commencent à monter en nombre à partir de 2016. Un point intéressant à noter est

que les études expérimentales ne sont pas le monopole des pays occidentaux. En fait, aucune étude expérimentale n'a été réalisée en Amérique du Nord. Nous en avons repéré en Afrique (3), en Europe (2) et en Asie (1). Dans le cas des études quasi expérimentales, la tendance est semblable puisqu'elles ont été réalisées principalement en Europe (25), en Afrique (16) et en Asie (9) (Tableau 17).

³⁷ Cette étude a été divisée en deux, car la cueillette des données, l'analyse et l'interprétation de données ont été différentes pour chaque pays.

Tableau 17. Design expérimental par continent

Continent	Expérimental		Quasi expérimental		Autres	
	n	%	n	%	n	%
Total	6	2,7 %	54	24,7 %	157	71,7 %
Afrique	3	6,0 %	16	32,7 %	30	61,2 %
Amérique du Nord	0	0,0 %	3	16,70 %	15	83,3 %
Asie	1	2,4 %	9	21,4 %	32	76,2 %
Europe	2	2,1 %	25	25,8 %	70	72,2 %
N/A	0	0,0 %	0	0,00 %	4	100,0 %
Océanie	0	0,0 %	1	14,30 %	6	85,7 %

On note la faible présence de designs expérimentaux (0 %) ou quasi expérimentaux (13,6 %) dans le domaine de la prévention tertiaire qui recourt donc presque exclusivement à d'autres types d'approche (86,4 %). Comme mentionné précédemment, cela peut s'expliquer en partie, particulièrement dans le cas de l'évaluation des effets, par les difficultés associées à la collecte de données dans ce domaine et aux enjeux éthiques de l'utilisation de modèles expérimentaux dans le cas de ce type de programmes de prévention. Les designs

expérimentaux et quasi expérimentaux se concentrent donc dans les programmes de prévention primaire et de prévention primaire ciblée ainsi que dans les programmes qui ciblent tout type d'extrémisme (Tableau 18). Dans la même veine, les designs quasi expérimentaux sont également présents dans le cas de programmes de prévention secondaire. À ces niveaux de prévention, l'utilisation de groupes témoins et de personnes participantes est en effet plus facile en raison d'un plus grand nombre d'individus.

Tableau 18. Design expérimental par niveau de prévention et type d'extrémisme

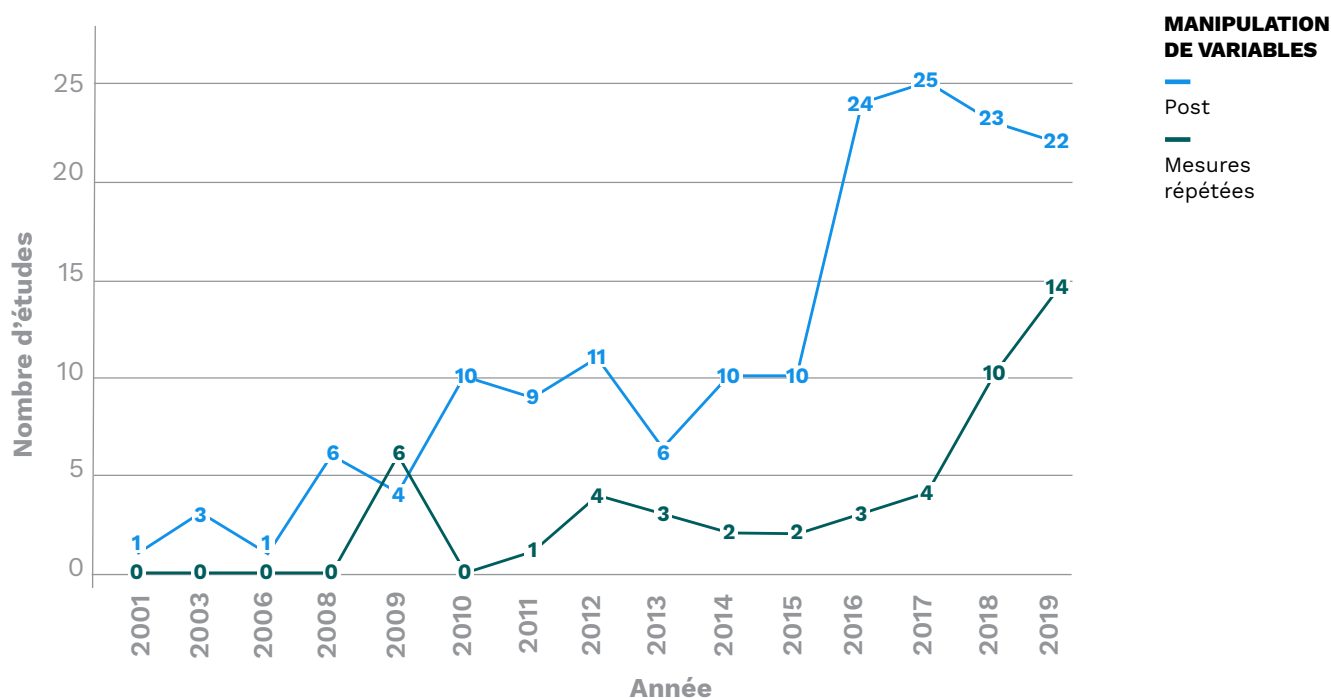
		Expérimental		Quasi expérimental		Autres	
		n	%	n	%	n	%
Niveau de prévention	Primaire	3	6,3 %	11	22,9 %	34	70,8 %
	Primaire ciblée	5	5,6 %	22	24,7 %	62	69,7 %
	Secondaire	0	0,0 %	16	26,7 %	44	73,3 %
	Tertiaire	0	0,0 %	6	13,6 %	38	86,4 %
	Générale	0	0,0 %	8	33,3 %	16	66,7 %
Type d'extrémisme	Extrême droite	1	5,0 %	5	25,0 %	14	70,0 %
	Islamisme	1	1,2 %	13	15,7 %	69	83,1 %
	Non spécifique	4	3,2 %	37	29,6 %	84	67,2 %

3.3.5 À quel moment les données ont-elles été recueillies ?

Les recensions de la littérature mentionnent souvent l'absence de designs à mesures répétées dans les évaluations de ces programmes, c'est-à-dire basés sur plusieurs observations des mêmes sujets à au moins deux moments différents. Dans le cas de cette étude, nous avons identifié les évaluations qui ont pris au moins une mesure avant et une après l'intervention, sans tenir compte du fait que la différence de temps entre la première et la dernière mesure était de quelques heures, jours ou mois.

Quarante-neuf études qui ont pris des mesures répétées ont été repérées (Tableau 19), dont 43 études reposent sur un design quasi expérimental et les six restantes, sur un modèle expérimental. Elles constituent environ un quart de toutes les études identifiées par cette revue systématique. Comme il est habituel dans ce type d'études, toutes ont cherché à tester l'efficacité des programmes en question et seulement 10,2 % ont également évalué le processus de mise en œuvre sur la base d'analyses qualitatives, c'est-à-dire qu'il s'agissait d'études mixtes.

Figure 15. Évolution des études à mesures répétées par année



Les modèles de prise de mesures répétées ont été présents dans le domaine depuis au moins 2009 et ont connu une augmentation importante dans les deux dernières années analysées dans cette étude (Figure 15). Ces résultats démontrent que ce type d'études commence

à devenir un standard dans le domaine et peut être perçu comme un signe de l'amélioration de la qualité des designs utilisés, notamment en ce qui concerne l'évaluation des effets des programmes.

Tableau 19. Présence de mesures répétées par continent, par niveau de prévention et par type d'extrémisme

		Mesures répétées			
		Non		Oui	
		n	%	n	%
Total		165	75,3 %	49	22,4 %
Continent	Afrique	39	79,6 %	10	20,4 %
	Amérique du Nord	16	88,9 %	2	11,1 %
	Asie	32	78,0 %	9	22,0 %
	Europe	69	71,9 %	27	28,1 %
	N/A	3	100,0 %	0	0,0 %
	Océanie	6	85,7 %	1	14,3 %
Niveau de prévention	Primaire	37	77,1 %	11	22,9 %
	Primaire ciblée	65	73,0 %	24	27,0 %
	Secondaire	46	78,0 %	13	22,0 %
	Tertiaire	35	85,4 %	6	14,6 %
	Générale	19	79,2 %	5	20,8 %
Type d'extrémisme	Extrême droite	11	64,7 %	6	35,3 %
	Islamisme	72	88,9 %	9	11,1 %
	Non spécifique	91	72,2 %	35	27,8 %

La majorité des études qui ont eu recours à des mesures répétées ont utilisé un design prépost simple ($n=37$), c'est-à-dire une prise de mesures (collecte de données) avant l'intervention et une autre après l'intervention. Six études supplémentaires se sont appuyées sur au moins une prise de mesure supplémentaire en plus de la prise de mesures prépost traditionnelle. Parmi ces six études, trois ont collecté des données entre la prépost intervention (*between mesures* ou *midline point*) et trois autres études ont effectué au moins un suivi supplémentaire (*follow-up*). C'est le cas, par exemple, de l'étude qui cherchait à examiner les effets à long terme d'un programme de formation visant à accroître la résilience contre l'extrémisme violent chez de jeunes vulnérables (DIAMOND, University of Amsterdam, 2013). L'évaluation a été réalisée à partir d'un modèle longitudinal dans lequel les évaluatrices et évaluateurs ont pris des mesures au cours de quatre moments (avant l'intervention, au milieu, immédiatement après et 3 mois après avoir terminé l'intervention). Il s'agit de l'un des rares exemples d'évaluation longitudinale à froid dans le domaine, c'est-à-dire ayant lieu un certain temps après l'intervention. Une dernière étude, organisée aux États-Unis par l'«*Homeland Security Department*», a utilisé la méthode des doubles différences³⁸ pour examiner si les événements d'engagement communautaire étaient associés à une réduction du contenu pro-ISIS sur Twitter (Mitts, 2017). L'auteur a recueilli les informations sur twitter avant les événements et 7, 14, 21 et 30 jours après ces derniers. Enfin, quatre études ont pris des mesures uniquement après la fin de l'intervention, mais ont effectué au moins un suivi supplémentaire quelque temps après. Finkel, Rojo-Mendoza, Schwartz, Belasco et Kreft (2015) ont, par exemple, évalué l'impact de l'une des composantes d'un programme de prévention au Niger et au Tchad, qui consistait à renforcer la capacité des stations de radio locales à produire et à diffuser des contenus de prévention de l'extrémisme violent élaborés par elles-mêmes. Le design méthodologique consistait en un panel expérimental longitudinal où un groupe sélectionné au hasard était incité à écouter les programmes radio. Ensuite, leurs données étaient comparées à celles d'un groupe témoin qui n'avait pas été encouragé. La collecte des données a eu lieu en deux vagues, une fois les messages radio diffusés et environ 10 mois plus tard.

Les études qui ont pris des mesures répétées sont particulièrement présentes dans le niveau de prévention primaire ciblée (27 %) et moins présente dans le niveau tertiaire (14,6 %). Ce dernier chiffre s'explique en raison des difficultés éthiques, méthodologiques et pratiques déjà mentionnées dans ce rapport. En revanche, les raisons de leur plus grande présence au niveau de la primaire ciblée ne sont pas tout à fait claires, considérant que ce niveau

ne présente pas nécessairement moins de difficultés que le niveau primaire. Les programmes de prévention de l'extrême droite sont aussi plus enclins à se faire évaluer avec des mesures répétées que les programmes qui ciblent l'islamisme radical. Ceci s'explique peut-être en raison du nombre plus limité des premiers où un peu de variation dans les chiffres peut amener à un changement important dans les pourcentages.

3.3.6 Savons-nous qui sont les participantes et participants qui ont pris part à ces études ?

La description détaillée des participantes et participants dans une étude est un standard dans la diffusion des études scientifiques et une norme de qualité pour les publications. Cependant, comme nous le voyons, cela n'a pas toujours été le cas dans les études évaluatives identifiées. Dans cette section, nous décrivons de façon globale ces personnes participantes ainsi que celles impliquées comme groupe témoin dans ces études. Le Tableau 20 est un résumé des échantillons utilisés par type de programme.

Sur le nombre total d'études identifiées, seulement 58,5 % ont indiqué le nombre de participantes et participants aux évaluations et seulement 10 % d'entre elles ont utilisé un groupe témoin, dont toutes les études d'évaluation d'impact utilisant des méthodes expérimentales (6) et quasi expérimentales avec un groupe témoin (16). L'absence d'une description relativement détaillée des personnes participantes dans tout modèle d'évaluation est un indicateur clair d'un manque de transparence méthodologique qui, dans le cas des études incluses dans cette revue systématique est, comme nous le verrons dans la section sur la qualité méthodologique, l'un des principaux défauts d'une partie considérable d'entre elles.

Ce manque de transparence est variable, selon des catégories utilisées. En Amérique du Nord, par exemple, la taille de l'échantillon n'est pas déclaré dans 61,1 % des études et, en Asie, dans 54,8 % des pourcentages qui sont clairement supérieurs au taux global (41,6 %). En revanche, cette information n'est pas présente dans un tiers des études en Océanie (28,6 %) et en Afrique (30 %). Lorsque l'on considère le niveau de prévention faisant l'objet d'une évaluation, l'information concernant l'échantillon n'est pas présente dans environ la moitié des évaluations des programmes de prévention primaire (52,1 %), tertiaire (47,8 %) et secondaire (47,5 %). Cette situation se répète également majoritairement dans les programmes de prévention de l'extrême droite (65 %).

³⁸ Cette analyse permet d'estimer l'effet d'une intervention en comparant la différence entre le groupe témoin et le groupe de traitement, avant et après l'introduction de ce dernier. Le modèle utilisé par Mitts (2017) est semblable à celui de séries temporelles.

Tableau 20. Nombre de participantes et participants par type d'étude

			n*	Total*	Moyenne	Écart type	Minimum	Maximum
Total		Personnes participantes	128	219	233,54	415,7	3	2789
		Groupe témoin	22		343,7	409	1	1452
Continent	Afrique	Personnes participantes	35	50	387,9	552,3	14	2789
		Groupe témoin	11		468,9	486,5	27	1452
	Amérique du Nord	Personnes participantes	7	18	199,3	269,7	26	767
		Groupe témoin	2		186,0	198,0	46	326
	Asie	Personnes participantes	19	42	358,9	563,3	4	1657
		Groupe témoin	3		227,0	189,6	25	401
	Europe	Personnes participantes	61	98	126,3	238,0	3	1113
		Groupe témoin	6		225,0	370,8	1	976
	N/A	Personnes participantes	1	4	154,0		154,0	154,0
		Groupe témoin						
	Océanie	Personnes participantes	5	7	49,2	39,2	16,0	117,0
		Groupe témoin						
Niveau de prévention	Primaire	Personnes participantes	23	48	512,0	653,9	9,0	2789,0
		Groupe témoin	6		710,5	544,3	102,0	1452,0
	Primaire ciblée	Personnes participantes	57	89	226,1	361,2	5,0	1657,0
		Groupe témoin	10		451,6	425,0	25,0	1050,0
	Secondaire	Personnes participantes	32	61	237,9	416,8	3,0	1644,0
		Groupe témoin	3		24,7	22,6	1,0	46,0
	Tertiaire	Personnes participantes	24	46	171,9	311,4	4,0	1170,0
		Groupe témoin	2		128,0	179,6	1,0	255,0
	Générale	Personnes participantes	15	24	116,4	105,5	4,0	368,0
		Groupe témoin	5		119,8	28,6	85,0	152,0
Type d'extrémisme	Extrême droite	Personnes participantes	20	151,9	269,1	3,0	747,0	20
		Groupe témoin		102,0		102,0	102,0	
	Islamisme	Personnes participantes	84	216,5	384,7	3,0	1644,0	84
		Groupe témoin		183,1	178,9	1,0	484,0	
	Non spécifique	Personnes participantes	127	245,9	461,1	4,0	2789,0	127

* : N = Nombre d'études qui mentionnent le nombre de participants ;

Total = Nombre total d'études identifiées

En revanche, toutes les études qui ont eu recours à des groupes témoins ont informé au moins quant au nombre de personnes participantes. L'utilisation d'un groupe témoin dépend du design choisi et des questions de recherche et elle est recommandée dans l'étude de l'efficacité d'un programme, ce qui, dans cette revue systématique, concerne principalement les expériences et les designs quasi expérimentaux. La présence ou l'absence relative des groupes témoin ne peut donc pas être utilisée comme un indicateur de la qualité de toutes les études retenues ici. Prenant en considération cette remarque, l'utilisation de groupes témoins est plus présente en Afrique (22 %) et en Amérique du Nord (11,1 %), dans les programmes de prévention primaire (12,5 %) et primaire ciblée (11,2 %) et dans les études qui évaluent des programmes généralistes (22,8 %) : autrement dit, dans les programmes plus universels, ce qui est

cohérent avec l'idée que ce type de programme est moins compliqué pour l'utilisation de designs quantitatifs plus sophistiqués.

La taille des échantillons est très variable entre les programmes allant de 3 à 2789 (écart type = 415,7) participantes et participants généraux aux programmes et de 1 à 1452 (écart type = 409) personnes participant aux groupes témoins. Cette variabilité implique que la moyenne est moins significative comme indicateur. À titre d'exemple, la médiane des personnes participantes est de 56, dans le cas des participantes et participants généraux, et de 148,5, celui des groupes témoins. Cela signifie que, dans le premier cas, la moitié des programmes qui ont mentionné la taille de leur échantillon ont fait appel à 56 participantes et participants ou moins dans le cadre du processus d'évaluation (Tableau 21).

Tableau 21. Nombre de participantes et participants par design global

		n*	Total*	Moyenne	Écart type	Minimum	Maximum
Quantitatif	Personnes participantes	21	30	376,3	419,7	16	1657
	Groupe témoin	10		429,7	429,8	25	1050
Qualitatif	Personnes participantes	52	96	48,3	66,5	3	357
	Groupe témoin	1		1,0		1	1
Mixte	Personnes participantes	55	91	354,1	530,3	5	2789
	Groupe témoin	11		296,6	404,3	46	1452

* : N = Nombre d'études qui mentionnent le nombre de participants ; Total = Nombre total d'études identifiées

Évidemment, le design global exerce une influence sur cette variabilité. Ainsi, les études qualitatives nécessitent moins de personnes participantes que les études quantitatives. Comme le montre le Tableau 21, la moyenne des études qualitatives est de 48,3 (médiane = 21,5) et la variabilité de l'échantillon demeure une question importante, des études ayant interrogé un minimum de 3 personnes et un maximum de 357. Malgré cette variabilité, la moitié des études qualitatives a porté sur 21 personnes ou plus. Ce type d'étude nécessite une diversité de points de vue et ces résultats peuvent être une preuve de la qualité des études qualitatives. En revanche, seulement 55 % des études qualitatives ont rapporté la taille de leur échantillon.

La moyenne des participantes et participants aux études quantitatives est de 376,3 (médiane=191) et 7 études quantitatives sur 10 ont mentionné la taille de leur échantillon. Par ailleurs, un tiers des études quantitatives ont eu recours à des groupes témoins.

Finalement, 60,4 % des études mixtes ont mentionné la taille de leur échantillon. Lorsqu'elles l'ont fait, la variabilité a été beaucoup plus significative que dans le cas des autres types d'étude et le recours aux groupes témoins, moins important. Comme on le verra plus tard, dans la section sur l'évaluation de la qualité méthodologique, ce type d'étude a été évalué comme étant de moindre qualité en ce qui concerne l'intégration des deux méthodologies.

3.3.7 Outils de collecte de données

La majorité des études a utilisé des moyens traditionnels de collecte de données, à savoir notamment des entretiens individuels (74 %), des sondages (49,8 %), des groupes de discussion (32,4 %) et, plus rarement, des observations (16,9 %) (Tableau 22). La catégorie « autres »

inclut, notamment, des données secondaires³⁹, tels que les registres ou la documentation des organismes impliqués dans l'évaluation ou des analyses de l'activité des réseaux sociaux en ligne.

Tableau 22. Outils de collecte de données par type d'extrémisme, par niveau de prévention et par objectif de l'évaluation⁴⁰

		Sondages		Entretiens		Groupe de discussion		Observation		Autres	
		n	%	n	%	n	%	n	%	n	%
Total		109	49,8 %	162	74,0 %	71	32,4 %	37	16,9 %	73	33,3 %
Niveau de prévention	Primaire	25	53,2 %	32	68,1 %	18	39,1 %	8	17,0 %	21	44,7 %
	Primaire ciblée	53	59,6 %	62	69,7 %	30	33,7 %	19	21,3 %	23	25,8 %
	Secondaire	33	55,0 %	48	80,0 %	23	38,3 %	6	10,0 %	27	45,0 %
	Tertiaire	14	31,8 %	39	86,7 %	10	22,7 %	5	11,1 %	18	40,9 %
	Générale	9	37,5 %	20	83,3 %	8	33,3 %	6	25,0 %	7	29,2 %
Type d'extrémisme	Extrême droite	8	42,1 %	11	57,9 %	0	0,0 %	4	21,1 %	5	26,3 %
	Islamisme	35	42,2 %	63	75,9 %	30	36,6 %	17	20,5 %	29	34,9 %
	Non spécifique	69	55,2 %	97	77,0 %	45	36,0 %	19	15,1 %	42	33,6 %
Objectif de l'évaluation	Impact	98	62,0 %	109	69,0 %	59	37,6 %	27	17,1 %	49	31,0 %
	Processus	44	40,4 %	103	93,6 %	40	36,7 %	22	20,0 %	33	30,3 %
	D'Extrants	13	50,0 %	23	88,5 %	17	65,4 %	4	15,4 %	19	73,1 %

Si l'on croise cette information avec d'autres catégories, on observe que les stratégies de cueillette de données changent en fonction du niveau de prévention, du type d'extrémisme et de l'objectif de l'évaluation.

Lorsque l'on considère les différents types d'extrémisme, on peut observer que les groupes de discussion constituent une stratégie importante dans le cas de l'évaluation des programmes de prévention destinés

à l'extrémisme islamiste et des programmes non spécifiques. En revanche, ce n'est pas le cas pour les programmes de prévention qui visent l'extrême droite. Pour ces programmes, la stratégie privilégiée est principalement celle de l'entretien individuel. Ceci peut s'expliquer, notamment, en raison du nombre limité de personnes participant à l'évaluation (la moitié de ces études a interrogé 15 personnes ou moins au cours de l'évaluation). Orban (2019), par exemple, explique

³⁹ Toutes ces données secondaires sont des sources d'information supplémentaires aux sources primaires et sont donc décrites ici même si nous n'avons retenu que les études qui ont utilisé des sources primaires.

⁴⁰ Seules les catégories principales ont été considérées dans cette analyse.

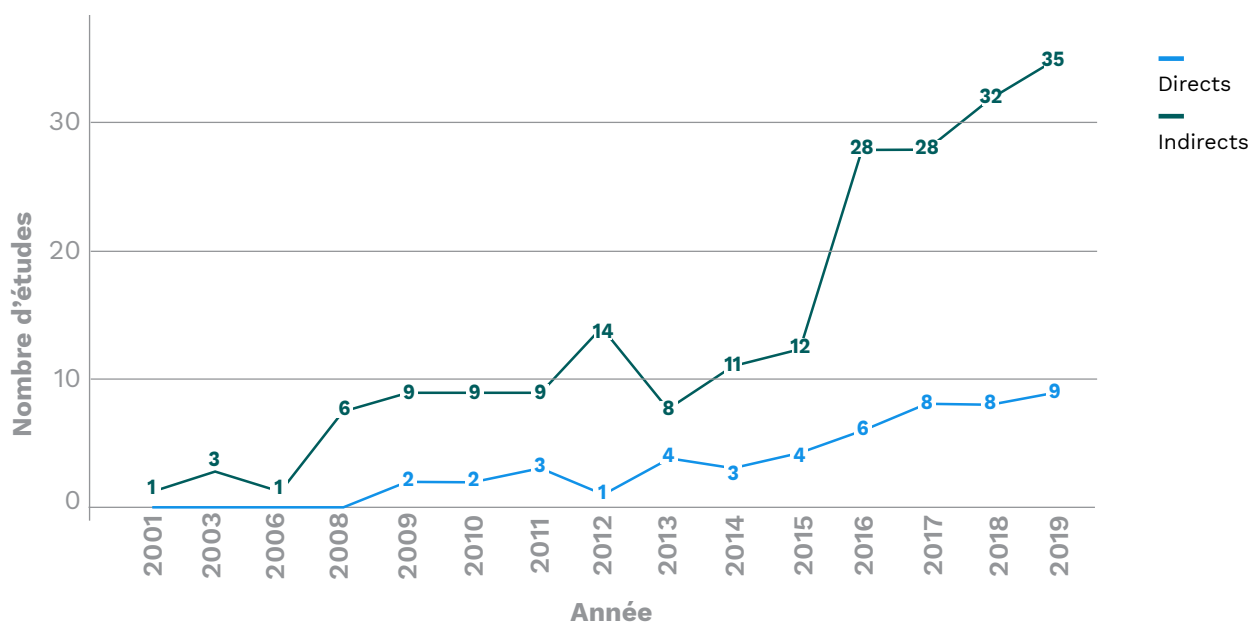
dans son étude que l'une des difficultés rencontrées, lors de l'évaluation d'un programme qui ciblait des individus d'extrême droite, a été de susciter la confiance des personnes interrogées. On observe une situation semblable dans le cas des évaluations de programmes en prévention tertiaire. Ceux-ci privilégient davantage les entretiens individuels que les évaluations des programmes de prévention primaire. Là encore, le faible nombre d'effectifs semble en cause.

De leur côté, les évaluations d'impact sont associées, comme il a été déjà mentionné, à une approche principalement quantitative. Cependant, dans le cas des études incluses dans cette revue systématique, on observe que les évaluations portant sur les effets de l'intervention ont utilisé une diversité d'options de cueillette de données. Cela tend à corroborer l'importance donnée aux méthodes mixtes et au point de vue des actrices et acteurs sur le terrain. Les entretiens sont présents dans 69 % des études tandis que les sondages ne sont présents que dans 62 % des études. Cela confirme l'idée que certaines évaluations ont mesuré l'impact en utilisant des approches qualitatives. Les évaluations de processus utilisent largement les outils de collecte qualitatifs tandis que les évaluations d'extrants ont eu majoritairement recours aux données secondaires (autres = 73,1 %).

3.3.8 Mesurer ou ne pas mesurer directement l'extrémisme violent : le type d'indicateur utilisé

Comme mentionné dans la section de l'introduction associée aux difficultés de l'évaluation, l'un des enjeux de ce type d'étude est l'utilisation d'indicateurs précis de performance ou de valorisation des actions entamées (Figure 16). Cela n'est pas toujours simple en raison, notamment, du flou qui entoure la définition de la radicalisation et de l'extrémisme violent. Afin de combler cette lacune, certains évaluateurs et évaluatrices recourent à des indicateurs indirects, c'est-à-dire des indicateurs qui ne mesurent pas directement la radicalisation, l'extrémisme violent ou les sympathies pour ces phénomènes. En revanche, ils se penchent sur des facteurs mesurables qui, théoriquement, sont associés à ce processus, tels que l'estime de soi, la résilience individuelle et communautaire ou la complexité intégrative. Cette section explore le recours à ces deux types d'indicateurs, directs et indirects, dans les études identifiées par la revue systématique.

Figure 16. Évolution de l'utilisation des indicateurs directs et indirects par année



La majorité des études a utilisé exclusivement des indicateurs indirects (74 %), alors que très peu d'entre elles ont recouru exclusivement à des indicateurs directs (4,1 %). Une étude sur cinq a utilisé les deux. L'utilisation d'indicateurs indirects a connu une augmentation considérable à partir de 2016. En revanche, bien que

l'utilisation d'indicateurs directs ait également connu une augmentation, celle-ci n'a pas été aussi importante que pour les indicateurs indirects (Figure 16). Ainsi, on ne trouve jamais plus d'une étude par année ayant utilisé exclusivement des indicateurs directs et cette tendance reste stable au cours des années. Cette information

tend à confirmer la difficulté à mesurer directement la radicalisation ou l'extrémisme violent et à avoir une approche pragmatique (voir section 1.1) afin d'évaluer ce type de programmes avec les ressources disponibles.

L'utilisation d'indicateurs selon les différentes régions est également intéressante. Dans les cas des indicateurs indirects, la tendance suit celle de l'analyse globale. En revanche, elle diffère de manière significative quant à l'utilisation d'indicateurs indirects ou directs (Tableau 23). Ces derniers sont davantage privilégiés en Afrique (40 %) et en Asie (28,6 %) par rapport aux autres régions. Deux raisons peuvent expliquer cette différence. La première est que ces régions, étant plus lourdement touchées par les actes terroristes, ont normalisé la présence de groupes terroristes et, par conséquent, les réponses aux questions directes sur l'extrémisme violent sont moins problématiques d'un point de vue éthique ou méthodologique. La deuxième explication est moins élégante. Comme mentionné auparavant, une grande partie de ces évaluations sont réalisées pour les bailleurs de fonds ou réalisées par des personnes ressortissantes des pays occidentaux, notamment dans le cas de l'Afrique. Nous pouvons donc envisager l'hypothèse que les chercheuses et chercheurs occidentaux ont moins de réticences à poser des questions directes sur la radicalisation ou l'extrémisme violent, questions plus difficiles à poser dans le contexte occidental.

Tableau 23. Type d'indicateur utilisé par continent

	Indirects		Directs	
	n	%	n	%
Afrique	48	96,0 %	20	40,0 %
Amérique du Nord	18	100,0 %	2	11,1 %
Asie	38	90,5 %	12	28,6 %
Europe	91	93,8 %	18	18,4 %
N/A	4	100,0 %	1	25,0 %
Océanie	7	100,0 %	1	14,3 %

Le Tableau 24 montre qu'il n'y a pas de différence majeure dans l'utilisation d'indicateurs en fonction du type d'extrémisme ou du type d'évaluation, sauf dans le cas des évaluations qui portent sur le processus de mise en place d'un programme. Dans ce dernier cas, les indicateurs directs sont moins utilisés, ce qui est conforme à la logique de l'évaluation.

Tableau 24. Type d'indicateur utilisé par niveau de prévention, par type d'extrémisme et par objectif de l'évaluation

		Indirects		Directs	
		n	%	n	%
Niveau de prévention	Primaire	46	95,8 %	16	33,3 %
	Primaire ciblée	86	96,6 %	17	19,1 %
	Secondaire	53	88,3 %	21	34,4 %
	Tertiaire	40	88,9 %	18	39,1 %
	Générale	22	91,7 %	4	16,7 %
Type d'extrémisme	Extrême Droite	17	85,0 %	6	30,0 %
	Islamisme	77	92,8 %	21	25,0 %
	Non spécifique	123	97,6 %	30	23,6 %
Type d'évaluation	Impact	149	93,7 %	48	30,2 %
	Processus	106	97,2 %	20	18,2 %
	D'extrants	26	96,3 %	7	25,9 %

Quant aux indicateurs, les tendances changent légèrement lorsque l'on considère le niveau de prévention. Les évaluations qui ciblent les niveaux les plus universels ont tendance à utiliser des indicateurs indirects dans presque la totalité des études. Ces indicateurs sont légèrement moins importants dans les cas des programmes plus spécialisés (secondaire et tertiaire). Les indicateurs directs sont aussi plus utilisés dans les cas de ces derniers ainsi que dans les programmes de prévention primaire. L'utilisation d'indicateurs directs dans le cas de l'évaluation des programmes spécialisés semble logique dans la mesure où c'est dans ce type de programme que l'on retrouve, en principe, des individus plus facilement classables, comme des personnes radicalisées ou à risque de se radicaliser. À l'inverse, l'utilisation d'indicateurs indirects dans des programmes plus généraux semble également être une option pertinente puisque le public cible des interventions n'est pas forcément associé aux groupes radicalisés. Cependant, d'un point de vue méthodologique, cette approche soulève des questions. En effet, les groupes associés aux niveaux secondaire et tertiaire sont probablement les plus politisés et les plus convaincus de leurs idées. Les interventions sont souvent effectuées dans un contexte d'obligation pénale (prisons) et hautement politisées. Dès lors, les réponses et les résultats liés aux indicateurs directs dans ce contexte comportent de sérieux biais. Madriaza, Valendru, Stock-Rabbat, Ponsot et Marion (2018), par exemple, ont indiqué avoir eu ce type de difficultés dans l'évaluation d'un programme au sein du système de probation français.

3.4 QUELS SONT LES LIMITES, LES DIFFICULTÉS ET LES CONFLITS D'INTÉRÊTS LIÉS À L'ÉVALUATION DES PROGRAMMES DE PRÉVENTION DE L'EXTRÉMISME VIOLENT ?

Toute étude qui utilise une méthode scientifique peut présenter des limites quant au design, à la collecte de données ou à l'interprétation des données. Le fait d'identifier les limites est considéré comme une bonne pratique en recherche puisque celles-ci nous informent des frontières au sein desquelles l'interprétation des données est valide, fiable et éventuellement transférable. À l'inverse, ne pas les évoquer laisse entendre que les recherches n'ont pas rencontré de difficultés au cours du processus de recherche scientifique ou que l'interprétation des données est valide sans égard au contexte de la recherche. La question des limites est d'autant plus essentielle que l'évaluation en prévention de l'extrémisme violent devrait avoir pour objectif de transférer les leçons apprises et d'améliorer les pratiques. Elle est indissociable de la question de la reproductibilité

d'un programme évalué positivement à un autre contexte de prévention.

Pour leur part, les conflits d'intérêts renvoient à la question de l'éthique de la recherche et aux facteurs externes ou internes au processus méthodologique qui peuvent avoir un impact sur les résultats. Au même titre que les limites, l'identification de ces conflits potentiels nous permet de mettre en perspective l'interprétation et la fiabilité des données. À la différence d'autres sections de ce rapport, nous analysons ici quantitativement et qualitativement ces deux éléments, autrement dit, nous décrivons également les principales limites, difficultés et conflits d'intérêts rapportés par les évaluatrices et évaluateurs (Tableau 25).

Tableau 25. Limites et conflits d'intérêts des évaluations

		Limites identifiées		Conflits d'intérêts identifiés		Conflits d'intérêts potentiels	
		n	%	n	%	n	%
Total		84	38,4 %	14	6,4 %	45	20,5 %
Type de publication	Littérature académique	28	31,5 %	4	4,5 %	23	25,8 %
	Littérature grise	56	43,1 %	10	7,8 %	22	16,9 %
Type d'évaluateur, d'évaluatrice	Interne	18	35,3 %	3	5,9 %	36	70,6 %
	Externe	64	40,5 %	10	6,4 %	9	5,7 %
Continent	Afrique	25	50,0 %	3	6,1 %	10	20,0 %
	Amérique du Nord	3	16,7 %	0	0,0 %	4	22,2 %
	Asie	21	50,0 %	4	9,5 %	8	19,0 %
	Europe	28	28,6 %	7	7,1 %	18	18,4 %
	N/A	3	75,0 %	0	0,0 %	2	50,0 %
	Océanie	4	57,1 %	0	0,0 %	3	42,9 %
Niveau de prévention	Primaire	16	33,3 %	2	4,3 %	9	18,8 %
	Primaire ciblée	28	31,5 %	5	5,6 %	22	24,7 %
	Secondaire	34	55,7 %	5	8,2 %	15	24,6 %
	Tertiaire	23	50,0 %	4	8,7 %	9	19,6 %
	Générale	5	20,8 %	3	12,5 %	4	16,7 %
Type d'extrémisme	Extrême droite	8	40,0 %	1	5,0 %	6	30,0 %
	Islamisme	37	44,0 %	4	4,8 %	22	26,2 %
	Non spécifique	42	33,1 %	8	6,3 %	20	15,7 %
Objectif de l'évaluation	Impact	65	40,9 %	10	6,3 %	39	24,5 %
	Processus	43	39,1 %	6	5,5 %	17	15,5 %
	D'extrants	12	44,4 %	2	7,4 %	5	18,5 %
Design global	Quantitative	14	46,7 %	3	10,0 %	8	26,7 %
	Qualitative	30	31,3 %	8	8,3 %	6	6,3 %
	Mixte	40	44,0 %	3	3,3 %	30	33,0 %

3.4.1 Études identifiant clairement les limites

À peine plus du tiers des études (38,4 %) abordent les limites de la recherche (Tableau 25). De manière surprenante, ce pourcentage est moins important dans le cas de la littérature académique (31,5 %) pourtant censée satisfaire des critères de publication plus stricts. La littérature grise aborde davantage les limites, mais cela reste moins d'un cas sur deux (43,1 %). La majorité des études, près des deux tiers, n'identifient donc pas clairement de limites, ce qui constitue un point négatif de l'évaluation méthodologique dans ce type d'étude. En effet, les difficultés inhérentes à l'évaluation évoquées dans l'introduction de ce rapport laissent penser que des limites importantes devraient être observées et clairement identifiées dans la majorité des cas. Dès lors, on peut penser que l'absence de limites identifiées dans les évaluations s'explique moins par l'absence de difficultés rencontrées que par le choix de ne pas les mentionner. Une hypothèse qui pourrait expliquer cette situation est la dépendance relative possible des évaluatrices et évaluateurs vis-à-vis de ceux qui commandent les programmes de prévention et des bailleurs de fonds. Cela peut alors se traduire par la pression de résultats positifs qui les pousserait à restreindre l'information en ce qui concerne ces difficultés pour ne pas nuire à l'impact de ces résultats. Cependant, en termes proportionnels, la faible différence entre les évaluations internes (35,3 %) et externes (40,5 %), ces dernières censées être moins influencées par ce type de facteurs, suggère que ce manque d'information est assez transversal dans le domaine.

L'identification des limites méthodologiques au sein des évaluations varie considérablement selon les zones géographiques. Environ une étude sur deux identifie les limites des évaluations des programmes de prévention en Océanie (57,1 %), en Afrique (50 %) et en Asie (50 %). En revanche, l'Amérique du Nord (16,7 %) et l'Europe (28,6 %), des régions qui ont une grande tradition en recherche et en évaluation, sont les régions où les limites des évaluations sont, de manière assez significative, les moins identifiées au sein des études. Ce constat peut sembler surprenant dans la mesure où nombre de difficultés liées à l'évaluation des programmes de prévention sont relevées par les chercheuses et chercheurs basés dans les pays occidentaux. Il résonne néanmoins avec le défi soulevé par plusieurs d'entre eux de conserver une évaluation indépendante des programmes.

La question des limites se pose également de manière différenciée selon les niveaux de prévention ciblés par les programmes évalués. La tendance observée est que plus le niveau de prévention est généraliste ou universel, moins les limites sont identifiées. Ainsi, les limites sont identifiées seulement dans 20,8 % des évaluations de programmes généralistes, dans 33,3 % des programmes de prévention primaire et dans 31,5 % des programmes de prévention primaire ciblée. En revanche, elles sont

mentionnées dans 55,7 % des programmes de prévention secondaire et dans la moitié des programmes de prévention tertiaire. Dans le cas de ces deux derniers, il est compréhensible que les auteures et auteurs des évaluations mentionnent davantage les limites de leur évaluation, étant donné que le champ d'action est plus circonscrit, que les effets recherchés sont mieux identifiés et que les risques associés à ces programmes sont plus importants. En revanche, compte tenu des effets négatifs de certains programmes qui ciblent spécifiquement une communauté, il est possible que les évaluations des programmes généralistes aient moins de difficultés au moment de collecter et d'analyser l'information que les évaluations des programmes spécifiques.

En ce qui concerne le design méthodologique, on note une différence significative. Les limites sont mentionnées par 46,7 % des études quantitatives. Elles sont identifiées seulement dans 31,3 % des études qualitatives. Cela peut sembler d'autant plus surprenant que ces dernières sont soumises à une contrainte évidente, à savoir que l'information qu'elles analysent dépend de la subjectivité des participantes et participants et de l'analyse.

Finalement, dans le cas de l'objectif de l'évaluation, les études qui ont une section dédiée à l'évaluation d'extrants informent légèrement davantage que les études qui ciblent l'impact et le processus du programme. Les évaluations d'extrants sont ainsi généralement plus conscientes des limites de ce type d'évaluation.

3.4.2 Conflits d'intérêts réels et potentiels

Les conflits d'intérêts n'ont été mentionnés que par 14 études au total, ce qui représente seulement 6,4 % du total des évaluations (Tableau 25). Ces études proviennent principalement de la littérature grise. Quatre d'entre elles évoquent l'existence de conflits d'intérêts, mais n'en précisent toutefois pas la nature exacte. L'étude de Khalil, Brown, Chant, Olowo et Wood (2019) mentionne qu'ils ont travaillé au sein d'un organisme ayant exercé une certaine influence sur le programme qu'ils évaluent. Sabir (2014) évoque, quant à lui, avoir été, dans le passé, arrêté et détenu à tort pour des faits de terrorisme lorsqu'il cherchait de l'information sur Internet, ce qui a été la motivation pour entamer l'évaluation du programme en question. En résumé, à de rares exceptions près, la question pourtant essentielle des conflits d'intérêts est en pratique presque évacuée de l'évaluation.

Nous avons donc également décidé d'analyser les conflits d'intérêts potentiels des études évaluatives qui n'ont pas été mentionnés par les auteures et auteurs. Cette analyse doit évidemment être considérée avec beaucoup de prudence dans la mesure où elle dépend des informations parcellaires disponibles dans les publications sélectionnées. En tout état de cause, nous avons identifié des conflits potentiels dans 20,5 % des

études retenues (Tableau 25). D'un point de vue plus analytique, les conflits d'intérêts potentiels peuvent être regroupés en trois catégories distinctes. La première catégorie regroupe les conflits pouvant exister lorsqu'au moins une des personnes évaluatrices du programme a également directement participé à son élaboration ou à sa mise en œuvre, ce qui représente 53,3 % des cas de conflits d'intérêt potentiels identifiés. La deuxième catégorie comprend les conflits qui peuvent exister lorsqu'au moins une des personnes évaluatrices du programme travaille dans l'organisme chargé de sa mise en œuvre sans nécessairement faire partie de l'équipe de conception et d'implantation de l'intervention (35,6 %). La troisième catégorie concerne l'origine du financement de l'évaluation : 6,7 % des études ont été financées par le même organisme qui a mis en place le programme.

Au lieu de décrire chaque pourcentage par catégorie, comme nous l'avons fait pour les conflits d'intérêts déclarés, il est plus utile d'examiner le rapport entre le nombre d'études dans lesquelles nous avons identifié des conflits d'intérêts et le nombre d'études qui ont ouvertement déclaré ces conflits. Dans l'ensemble, nous avons identifié 3,2 études présentant des conflits d'intérêts pour chaque étude les ayant ouvertement déclarés. Ce ratio varie fortement en fonction de la catégorie de comparaison. Par exemple, pour chaque évaluation interne ayant déclaré un conflit d'intérêts potentiel, nous avons trouvé 12 études susceptibles de présenter ce

type de conflit, ce qui est d'autant plus inquiétant que c'est précisément ce type d'étude qui devrait être le plus attentif à ces questions. En revanche, nous avons identifié moins d'évaluations externes présentant des conflits d'intérêts potentiels que les études qui en ont fait état dans leurs publications. Un autre cas extrême est celui des études mixtes dans lesquelles, pour chaque étude ayant déclaré des conflits d'intérêts, nous en avons identifié 10. En retour, nous avons identifié moins d'études purement qualitatives avec des conflits potentiels que celles qui ont été ouvertement déclarées. Moins extrêmes, mais tout aussi inquiétants sont les programmes de prévention de l'extrémisme violent de droite (ratio = 6), de l'extrémisme islamiste (ratio = 5,5) et les publications académiques (ratio = 5,8). Ces dernières sont surprenantes compte tenu de l'obligation qu'ont de nombreuses revues académiques de faire état de ces conflits.

3.4.3 Limites identifiées par les auteures et auteurs

Jusqu'à présent, nous avons évoqué l'absence de limites identifiées dans la majorité des études. Toutefois, l'analyse du contenu des limites rapportées par 84 études offre une perspective enrichissante sur les difficultés rencontrées par les évaluatrices et évaluateurs sur le terrain. Nous avons classé ces limites en cinq catégories associées à la méthodologie, à l'analyse, à la personne évaluatrice, au programme et aux facteurs externes (Encadré 4).

Encadré 4. Limites identifiées par les études

1. Méthodologiques

a. Design

- i. Design exploratoire des études
- ii. Absence de prises de mesures avant et après l'intervention

b. Indicateurs

- i. Limités
- ii. Inadéquats

c. Cueillette de données

- i. Données basées sur la perception
- ii. Crédibilité de l'information obtenue
- iii. Outils de collecte de données non adéquats ou limités
- iv. Accès aux données
- v. Information incomplète

d. Personnes participantes

- i. Petit échantillon
- ii. Échantillon non représentatif
- iii. Manque de groupes témoins
- iv. Échantillon homogène
- v. Biais de désirabilité sociale
- vi. Réticences des participantes et participants

2. Analytiques

- a. Capacité d'analyse
- b. Profondeur des analyses
- c. Montée en généralité
- d. Type d'analyse réalisé
- e. Type de résultat obtenu
- f. Relations causales
- g. Informations sensibles

3. Évaluatrices et évaluateurs

- a. Indépendance de la personne évaluatrice
- b. Jugement subjectif de l'analyse
- c. Problèmes de traduction

4. Programme

- a. Concernant le programme
 - i. Courte période du déroulement du programme
 - ii. Manque de définitions claires
 - iii. Financement du programme
- b. Concernant les intervenantes et intervenants
 - i. Engagement
 - ii. Sécurité
 - iii. Rôle des personnes intervenantes externes

5. Externes

- a. Limites de temps
- b. Limites contextuelles et de sécurité
- c. Limites budgétaires

a. Les limites méthodologiques

Les limites méthodologiques sont celles les plus fréquemment mentionnées par les évaluatrices et évaluateurs. Elles affectent la plupart des étapes du processus d'évaluation : le design, les indicateurs, la collecte de données et les personnes qui participent à l'étude. Sur le plan du *design*, les évaluatrices et évaluateurs mentionnent l'approche exploratoire des études qui conduit à des analyses excessivement générales et qui empêche de tirer des conclusions sur l'efficacité des programmes évalués. Par exemple, c'est le cas de l'étude de Glazzard et Reed (2018) qui ont tiré leurs conclusions à partir d'un échantillon de programmes locaux en Europe sans engager un travail exhaustif sur le terrain et sur chaque programme. L'évaluation de l'impact est également limitée par un design qui ne considère pas la prise de mesures avant et après l'intervention. Cela s'explique souvent, comme dans le cas de l'étude de Johns, Grossman et McDonald (2014), par le fait que l'évaluation est initiée à mi-parcours du programme.

Les *indicateurs* utilisés limitent aussi les analyses lorsqu'ils restreignent les résultats à un seul aspect de la réalité. Garaigordobil (2012), par exemple, considère que l'approche psychologique et centrée sur l'analyse individuelle qu'elle a utilisée ne permet pas d'expliquer complètement le phénomène en raison de variables relationnelles et psychosociales qui sont également importantes dans le processus d'évaluation. Dans d'autres cas, les indicateurs n'ont pas été adéquats pour évaluer l'un des objectifs du programme. Van der Heide et Schuurman (2018) indiquent ainsi avoir dû utiliser le taux de récidive comme indicateur de déradicalisation. De leur côté, McDowell-Smith, Speckhard et Yayla (2017) mentionnent qu'ils se sont principalement concentrés sur un aspect d'une campagne de contre-discours (l'augmentation ou la diminution des attitudes négatives existantes à l'égard de Daesh) au lieu de considérer un point de vue plus large (l'accroissement de l'intérêt pour ce groupe, par exemple).

La phase de *cueillette de données* est également essentielle dans le processus d'évaluation. Les évaluatrices et évaluateurs mentionnent plusieurs défis et difficultés rencontrés pendant cette phase qui peuvent limiter l'interprétation de l'information : a) le type de données collectées, b) la crédibilité de l'information obtenue, c) l'outil de collecte, d) l'accès à l'information et e) l'information incomplète.

a) Le type de données collectées a évidemment une influence considérable sur les analyses. C'est le cas, notamment, des éléments de déclaration tels que les entretiens, les documents internes et les sondages de perception, qui reflètent davantage, d'après certains chercheurs et chercheuses, la perception des personnes intervenantes, conceptrices et usagères du programme

que la « réalité opérationnelle » au jour le jour (Gatewood et Boyer, 2019 ; Hirschi et Widmer, 2012 ; Rodon, 2018 ; Wilson et Krentel, 2018). La forte subjectivité de ces perceptions soulève la question de savoir dans quelle mesure il est possible de se fier à ces données.

b) La crédibilité de l'information obtenue est une limite évidente dans ce contexte. Les biais pouvant teinter l'information en question sont multiples : le caractère sensible du sujet sur le plan politique et de la sécurité (Sarota, 2017), les données collectées à partir du souvenir des personnes interviewées (Khalil et al., 2019) et parfois des personnes chercheuses elles-mêmes (Tsuroyya, 2017), des informations contradictoires entre les personnes interviewées (Cockayne et al., 2015) ou les documents consultés (Vittum et al., 2016).

c) Les résultats peuvent être également limités par les outils de collecte de données. C'est notamment le cas des études qui ont dû se limiter à un nombre restreint de questions dans les sondages (Parker et Lindekilde, 2020) ou qui ont été limitées par les biais de traduction, particulièrement dans des contextes non anglophones (Murtaza et al., 2018).

d) **L'accès à l'information est probablement le problème le plus récurrent rencontré par les évaluatrices et évaluateurs sur le terrain.** De nombreux chercheurs et chercheuses mentionnent, par exemple, l'impossibilité d'avoir accès à des informateurs clés tels que les personnes usagères du programme (Anindya, 2019 ; Jailobaeva et Asilbekova, 2017 ; Murtaza et al., 2018 ; Tines et al., 2017), une faction paramilitaire (Dwyer et Maruna, 2011), les membres du gouvernement (Jailobaeva et Asilbekova, 2017) ou les leaders religieux ou de la communauté (Muncy et al., 2015 ; Sarota, 2017 ; Wilson et Krentel, 2018). À titre d'exemple, Harahap, Irmayani et Lubis (2019) n'ont pas pu réaliser d'entretiens auprès de certains élèves dont les parents ont été des terroristes connus parce que c'était considéré comme un tabou dans la communauté. Il en résulte que la plupart de ces évaluations n'ont pris en considération que le point de vue des personnes intervenantes et conceptrices des programmes. Si, pour certaines et certains, cette source est suffisante, d'autres la considèrent comme incomplète, car elle dépend principalement de la perception des personnes directement impliquées dans le projet. Wilson et Krentel mentionnent, par exemple, leur incapacité à recueillir des données auprès des communautés : « L'équipe d'évaluation n'a pas été en mesure de vérifier les données fournies par les praticiens sur leurs activités de projet par le biais d'observations ou d'entretiens de suivi » (2018, p. 4). Certains évaluateurs et évaluatrices mentionnent les raisons qui expliquent les problèmes d'accès à des informateurs clés. Ils évoquent notamment la courte période allouée à l'évaluation et le problème de *timing* entre la période d'évaluation

et la disponibilité de certains acteurs et actrices sur le terrain. Dans des régions où il y a des situations de conflits, le contexte d'insécurité ou les limitations d'accès imposées par le gouvernement constituent un obstacle au bon déroulement de l'évaluation. Finalement, d'autres chercheurs et chercheuses mentionnent des problèmes logistiques tels que le changement de numéro de cellulaire des participantes et participants ou les périodes de congés.

Dans d'autres situations, la documentation des programmes est manquante ou incomplète en raison de l'incapacité de certains programmes et organisations à systématiser l'information qu'ils produisent (Ipp et al., 2014; Muncy et al., 2015; Rodon, 2018; Williams et al., 2016). Rodon explique ainsi :

L'analyse a mis à jour les limitations des données récoltées du fait des procédures pour leur renseignement. Ceci illustre bien combien le travail d'évaluation est largement dépendant de la rigueur du processus d'un dispositif d'intervention concernant ce qui peut véritablement être établi comme résultats ou non. Autrement dit, le niveau de professionnalisation des interventions notamment dans les procédures et méthodes de renseignement conditionne la portée des résultats et conclusions qui peuvent être établis à l'issue d'une évaluation d'impact. (2018, p. 111).

À ce problème de systématisation s'ajoute l'absence de données de référence, notamment au début du programme, qui empêche la comparaison des données recueillies (Ipp et al., 2014; Schumicky-Logan, 2017; Swedberg et Reisman, 2013). Dans certains cas, les chercheuses et chercheurs ont dû se fier à des souvenirs des actrices et acteurs du terrain pour reconstruire ces données de référence. Cela introduit des biais évidents pour l'évaluation dans la mesure où ces données ne peuvent pas être vérifiées de façon indépendante (Ipp et al., 2014). Les programmes ne sont évidemment pas conçus pour répondre aux besoins de la recherche et de l'évaluation, mais pour mettre en place des actions qui permettent la prévention de l'extrémisme violent. Cependant, le manque de sensibilité de certaines personnes actrices du terrain aux enjeux de l'évaluation peut avoir une influence significative sur l'évaluation de ces actions. Un constat qui confirme l'importance d'inclure, dès la conception du programme, une équipe d'évaluateurs et d'évaluatrices et de sensibiliser les personnes intervenantes sur ces enjeux.

e) Un dernier enjeu associé à la cueillette de données réside dans l'information incomplète au cours de ce processus. Cet enjeu concerne, notamment, le taux

de réponse et la perte d'information. Le faible taux de réponse semble être un problème assez récurrent dans ce domaine, notamment en lien avec les sondages (Beider et Briggs, 2010; Hiariej et al., 2017; Hirschi et Widmer, 2012; Schorn et al., 2010; Tines et al., 2017). Par exemple, Hirschi et Widmer (2012) mentionnent que dans l'un des programmes de prévention de l'extrême droite qu'ils ont évalués en Suisse, seules 4 personnes enseignantes sur les 107 contactées ont rempli le questionnaire. De son côté, l'équipe de Schorn (2010) explique le faible taux de réponse en raison de la « fatigue du sondage », les responsables de la mise en œuvre du programme venant de terminer l'évaluation de l'impact de celui-ci. La plupart de ces études ont utilisé des sondages à distance (en ligne, par téléphone ou par courriel), ce qui peut expliquer le faible taux de réponse. Les chercheuses et chercheurs mentionnent également le désintérêt pour le sujet, le manque de temps et les problèmes d'accès au terrain déjà évoqués. Quant à la perte d'information, elle concerne principalement la suppression de comptes sur les médias sociaux dans le cas de programmes en ligne (l'Encadré 5 explore cette question de manière plus approfondie).

Finalement, les limites associées aux participantes et participants sont également nombreuses dans ces études. La limite la plus souvent évoquée est la taille de l'échantillon, souvent petit pour le type d'analyse et les résultats attendus⁴¹. Comme nous l'avons vu dans la section consacrée aux échantillons, il y a des études qui ne mentionnent pas le nombre de participantes et participants à l'évaluation tandis que d'autres évaluent un programme en ayant recours à trois personnes participantes⁴². La représentativité, notamment en ce qui concerne les études quantitatives, constitue aussi un enjeu important pour de nombreuses études indépendamment de la taille de l'échantillon. En fait, il est rare de trouver des études représentatives et la plupart reposent sur des échantillons non probabilistes sélectionnés pour leur facilité d'accès ou déterminés par la motivation personnelle des participantes et participants à s'y impliquer. Cela conduit à certains biais dans la sélection des personnes participantes, qui peuvent entraver l'interprétation des données, en particulier la capacité à établir des liens de causalité entre les actions du programme et leurs résultats. L'absence de groupe de comparaison est également souvent mentionnée comme une limite pour établir ces liens de causalité. Seules 22 études ont eu recours à un groupe témoin. Il en résulte des études exploratoires, générales et descriptives qui peuvent difficilement évaluer les effets des programmes. Dans le cas des études qualitatives, il s'agit moins d'une question de représentativité que d'homogénéité ou de manque de diversité au sein des actrices et acteurs

⁴¹ Awan, 2012b; Busher et al., 2017; Cherney et Belton, 2019; Gatewood et Boyer, 2019; Hirschi et Widmer, 2012; Joyce, 2018; Kollmorgen et al., 2019; Kollmorgen et Barry, 2017; Madriaza et al., 2018; Mansour, 2017; Octavia et Wahyuni, 2014; Schorn, Moubayed, et Auten, 2010; Sjøen et Mattsson, 2019; Tines et al., 2017; Warrington, 2018.

⁴² La moitié des études qui informent leur échantillon ont moins de 56 participantes et participants.

interrogés⁴³. Certaines personnes actrices clés sont manquantes ou les personnes interviewées n'ont pas forcément le même profil que la population que les intervenantes et intervenants essayent de rejoindre.

Au-delà des problèmes d'échantillonnage, les chercheuses et chercheurs mentionnent une série de difficultés associées plus particulièrement aux motivations des personnes participantes. C'est notamment le cas du « biais de désirabilité sociale » qui consiste à vouloir se présenter sous un angle favorable à ses interlocuteurs et interlocutrices et à répondre conformément à leurs souhaits (Azam et Bareeha, 2017 ; Hiariej et al., 2017 ; Johns et al., 2014 ; Khalil et Ipp, 2016 ; Kollmorgen et al., 2019 ; Kollmorgen et Barry, 2017). Azam et Bareeha (2017) mentionnent, par exemple, que dans le cas de l'évaluation d'un programme de réhabilitation, la réponse positive des personnes participantes concernant le programme pourrait refléter leur intérêt à être considérées comme réhabilitées. Pour leur part, Hiariej et ses collègues (2017) se sont moins fiés aux réponses des responsables gouvernementaux impliqués dans ce processus d'évaluation, car ils ont eu tendance à fournir des réponses socialement souhaitables et probablement politiquement neutres. Lors de l'évaluation d'un programme pour l'emploi, les évaluatrices et évaluateurs signalent la possibilité d'une « sous-déclaration dans l'espoir de recevoir une formation ou des ressources supplémentaires, ou une surdéclaration pour valider les désirs perçus par Mercy Corps » (Kurtz, 2015, p. 15). Ces situations soulèvent la question de la confiance envers l'information fournie par les participantes et participants. Cependant, le cas contraire est tout aussi important. Il est essentiel d'obtenir la confiance des personnes participantes afin qu'elles se sentent libres de partager leurs impressions et d'éviter ainsi les réticences naturelles dans un domaine complexe. Le même Kurtz mentionne cet enjeu par rapport aux évaluatrices et évaluateurs étrangers : « nous reconnaissons la possibilité de biais dans les réponses, car les répondants peuvent se méfier de la raison pour laquelle les données ont été recueillies, surtout si elles sont recueillies par un étranger » (2015, p. 15). La réticence des actrices et acteurs à participer à ces évaluations peut aussi être le résultat de l'action ou du manque d'action des programmes eux-mêmes. Par exemple, Pipe, Egal, Malla, Billow et Abdi (2016) ont rencontré l'hostilité des membres de la communauté, car une seule des cinq actions prévues dans un programme avait été mise en œuvre au moment de l'évaluation. Finalement, la peur des personnes participantes à l'égard des représailles des factions gouvernementales et de l'opposition peut également freiner certains individus à participer à ces évaluations dans un contexte de violence et de conflit accru (Kurtz, 2015).

b. Les limites analytiques

Une fois la cueillette de données achevée, certaines limites peuvent également marquer la phase d'analyse des résultats. Une de ces limites est associée à la *capacité d'analyse* des chercheuses et chercheurs qui ont collecté des masses importantes d'information, notamment dans le contexte des devis qualitatifs. Par exemple, Dietrich (2018) a évalué le projet « White Dove » au Nigéria, qui a géré un centre de messagerie contre l'extrémisme violent composé de trois programmes radio. Au total, 824 personnes ont participé à des entretiens directs ou à des groupes de discussion. En addition à cette information, l'évaluation a documenté les réactions des auditrices et auditeurs des 44 premiers épisodes. Cette masse d'information a fait en sorte que les analyses n'ont pas pu prendre en considération une partie de l'information collectée, notamment les réactions des personnes auditrices.

Dans d'autres cas, il s'agit de la *profondeur des analyses* qui demeure parfois un survol global et descriptif du programme et de ses capacités. Nicolls et Hassan regrettent, par exemple, qu'« il n'a pas été possible de découvrir tous les disparités, facteurs et influences sur la mise en œuvre et les résultats » (2014, p. 50). Les désagréments des données en catégories plus spécifiques d'information sont, par conséquent, très limitées. Le problème d'analyse est également associé à la *tendance à la généralisation* à partir de résultats qui sont bornés d'abord par les difficultés de collecte de données mentionnées précédemment (faible nombre de personnes participantes, accès aux données, temps alloué à l'évaluation, etc.). Les résultats et les apprentissages obtenus ne concernent ainsi que le programme spécifique, les personnes qui ont participé à l'évaluation ou les secteurs géographiques où le programme est situé. C'est une situation qui limite à son tour la capacité de la politique publique à tirer des conclusions sur les programmes mis en place et, évidemment, sa transposition vers d'autres secteurs de la société ou du pays. Le *type d'analyse* réalisé a aussi une influence sur la généralisation de l'information. Williams et ses collègues (2016) mentionnent, par exemple, que l'utilisation des statistiques inférentielles/probabilistes dans son étude contient un élément d'incertitude quant à sa capacité de généraliser les résultats au-delà du programme lui-même. Dans d'autres cas, ce problème est lié aux *résultats* eux-mêmes qui, bien qu'ils puissent être statistiquement significatifs, ont une faible taille d'effet (Parker et Lindekilde, 2020). La taille de l'effet nous permet de déterminer, dans le contexte d'une évaluation d'impact causale, la force de l'influence de l'intervention sur les résultats attendus. Il est donc question ici, et dans plusieurs autres évaluations, de la

⁴³ Anindya, 2019 ; Awan, 2012b ; Azam et Bareeha, 2017 ; Bala et Deman, 2017 ; Busher et al., 2017 ; Hirschi et Widmer, 2012 ; Kollmorgen et al., 2019 ; Mansour, 2017 ; Muncy et al., 2015 ; Sarota, 2017 ; Schorn et al., 2010 ; Sjøen et Mattsson, 2019 ; Tesfaye, McDougal, Maclin et Blum, 2018 ; Warrington, 2018.

capacité des chercheuses et chercheurs, dans le contexte de leurs analyses, à établir une *relation causale* entre les programmes et ses résultats (Cherney et Belton, 2019 ; Demant, Wagenaar, et van Donselaar, 2009 ; Education Development Center (EDC) et USAID, 2019 ; Wilson et Krentel, 2018). Finalement, les analyses présentent des limitations en raison des caractéristiques inhérentes au sujet à l'étude. Cherney et Belton (2019) mentionnent, par exemple, qu'en raison des *informations sensibles* qui ont été collectées, seule l'une des personnes évaluatrices a codé les données, ce qui augmente le degré de subjectivité des analyses réalisées.

c. Les limites associées aux évaluatrices et évaluateurs

Trois limites associées aux évaluatrices et évaluateurs se répètent dans les articles retenus : l'indépendance de la personne évaluatrice par rapport au programme, le jugement subjectif de l'analyse et les problèmes de traduction. Dans tous les cas, il s'agit de l'inclusion de certains biais dans l'analyse qui peuvent donner comme résultat un portrait erroné des résultats obtenus. La question de l'*indépendance* et de la distance nécessaire pour évaluer un programme est souvent posée (Broadbent, 2013 ; Joyce, 2018 ; Rodon, 2018 ; Wilson et Krentel, 2018). Il n'est pas rare que des équipes de recherche aient été impliquées depuis le début dans l'élaboration du programme évalué (Broadbent, 2013 ; Madriaza et al., 2018b ; Wilson et Krentel, 2018). Une pratique qui est, en fait, recommandée par les praticiennes et praticiens et par les chercheuses et chercheurs comme un moyen de combler les distances entre les équipes et le manque de connaissance du terrain des évaluatrices et évaluateurs externes (Madriaza et al., 2021). Cependant, cette implication peut pousser les personnes évaluatrices à avoir un avis positif du programme et à négliger les aspects à améliorer. Dans le même ordre d'idées, et notamment dans le cas des études qualitatives, les évaluatrices et évaluateurs mentionnent les limites de l'analyse basée sur leur *propre jugement*, car elle reflète notamment leur point de vue. Azam et Bareeha le disent de façon plus claire : « L'analyse et les outils utilisés étant de nature qualitative, l'étude est basée sur les propres jugements et analyses de l'auteur, qui est ouvert à l'examen » (2017, p. 7).

Les problèmes de traduction ont déjà été évoqués dans la section concernant les limites méthodologiques. Il s'agit, notamment, du cas des évaluatrices et évaluateurs étrangers qui réalisent des évaluations dans des pays où ils ne parlent pas la langue et ont eu besoin d'avoir recours à des personnes traductrices ou interlocutrices pour réaliser les entretiens sur le terrain (Dhungana et al., 2016 ; Kollmorgen et Barry, 2017 ; Murtaza et al., 2018 ; Swedberg et Reisman, 2013). Cette limite est plus compliquée parce qu'il s'agit de chercheuses et chercheurs qui n'ont pas eu un accès direct au terrain, qui ont dû utiliser les réflexions des intermédiaires et qui n'ont pas pu lire la documentation concernant le

programme écrite dans une autre langue que la leur, laquelle est le plus souvent l'anglais. Cette situation se répète particulièrement dans les programmes financés par l'Agence de développement international américain (USAID). Si celle-ci a été une source très riche d'information pour cette revue systématique, cela soulève des questionnements concernant l'utilité de ce type d'évaluation et la fiabilité des données analysées pour les pays ciblés par l'agence qui se trouvent, notamment, en Afrique, au Moyen-Orient et en Asie centrale et Sud.

Finalement, en dehors de ces trois principales catégories, deux autres limites sont mentionnées dans les articles retenus : le biais de genre et les problèmes de santé des chercheuses et chercheurs. Le biais de genre est étonnamment absent des discussions des personnes évaluatrices. Ceci peut être en raison de la *parité relative* parmi les auteures et auteurs des publications. Seules Kollmorgen et Barry (2017) mentionnent ce biais, mais pour souligner l'absence d'hommes dans son équipe d'évaluation, le seul coéquipier ayant abandonné au début du processus d'évaluation. Quant aux problèmes de santé, Bean, Hill, Sany, et Riveles (2011) mentionnent que le temps alloué à l'évaluation a été perturbé par les problèmes de santé rencontrés par l'équipe de terrain.

d. Les limites concernant le programme

Une autre catégorie de limites se traduit, notamment, par les difficultés rencontrées par les équipes en lien avec le programme évalué et les intervenantes et intervenants. Les personnes évaluatrices invoquent souvent le fait que *la courte période du déroulement d'un programme* les empêche de tirer des conclusions plus décisives sur l'efficacité de l'action (Demant et al., 2009 ; Dhungana et al., 2016 ; Jailobaeva et Asilbekova, 2017). C'est le cas, notamment, du programme de prévention de l'extrême droite évalué par l'équipe de Demant (2009), qui a duré moins d'un an, car il a été considéré comme un pilote à l'intérieur de la stratégie municipale à la ville de Winschoten, aux Pays-Bas. Comme évoqué précédemment, il arrive souvent que certains programmes ne documentent pas leurs actions de manière systématique et que *certaines définitions ne soient pas suffisamment claires* pour déduire des conclusions plus approfondies. Bastug et Evlek (2016) ont ainsi évalué un programme pilote de désengagement et de déradicalisation individuelle en Turquie. Toutefois, ce dernier n'établissait pas de distinction claire entre ce qui était fait pour le désengagement et ce qui était fait pour la déradicalisation, et les deux mesures se chevauchaient fréquemment. Le *financement du programme* peut aussi avoir une influence sur l'évaluation. Dans un programme de formation au sein des écoles en Australie, l'équipe de Harris-Hogan (2019) n'a pas pu analyser les résultats de certaines juridictions dans le pays parce qu'il n'existait pas de financement central de la part du gouvernement fédéral pour mettre en place ces actions.

Quant aux personnes intervenantes, trois problèmes ont été soulevés en lien avec les limites mentionnées par les auteures et auteurs : l'engagement, la sécurité des personnes intervenantes et la participation des intervenantes et intervenants externes au programme d'intervention. *L'engagement* a été notamment mentionné par Frenett et Dow (2015) dans le cas d'un programme d'intervention directe en ligne (Encadré 5). Celui-ci était associé au fait que les intervenantes et intervenants avaient déjà des emplois à temps plein, qu'ils participaient à d'autres projets semblables, que l'intervention directe demandait beaucoup de temps et était épuisante et qu'ils recevaient un montant minimal comme rétribution à leur implication. Un autre enjeu mentionné par ces auteures et auteurs est la *sécurité des*

intervenantes et intervenants. L'intervention impliquait les échanges directs sur Facebook avec des personnes considérées comme des extrémistes, ce qui a soulevé la préoccupation des personnes intervenantes quant à leur sécurité, sujet fort important dans le domaine et qui a pourtant été négligé par la littérature (Madriaza et al., 2017). Le dernier point a été évoqué par Schuurman et Bakker (2016) dans le cas d'un programme de réinsertion aux Pays-Bas. *L'externalisation de certaines interventions* dans le programme hollandais envers des consultantes et consultants privés a limité la capacité des chercheuses et chercheurs à tirer des conclusions concernant le programme parce qu'ils ne connaissaient pas l'influence que ces personnes intervenantes ont eue sur les bénéficiaires du service.

Encadré 5. Difficultés de l'évaluation des programmes en ligne

L'évaluation des programmes en ligne soulève des enjeux assez différents de ceux des autres programmes. C'est le cas, notamment, de l'évaluation réalisée par Amanullah et Harrasy (2017) concernant une campagne soutenue par l'« Institute for strategic dialogue » (ISD), au Kenya. Ils ont développé 18 vidéos de contre-discours pour combattre le recrutement et la propagande extrémistes en utilisant les plateformes de médias sociaux. Au cours de la diffusion de ces vidéos, un compte Twitter utilisé pour la diffusion anonyme de la campagne a été inopinément bloqué par Twitter, ce qui a limité la capacité de l'équipe d'évaluation à recueillir des données pour certaines des vidéos.

Dans un autre projet d'ISD qui visait l'intervention directe sur des personnes qui expriment ouvertement des propos extrémistes en ligne sur Facebook, les chercheuses et chercheurs ont rencontré une diversité de problèmes associés aux nouvelles technologies (Frenett et Dow, 2015). D'abord, l'outil de recherches des profils sur Facebook (Graph Search) est devenu de plus en plus restreint en ralentissant le rythme auquel les candidates et candidats pouvaient être identifiés. Le projet a également été affecté par la suppression de profils par Facebook. Sur les 154 profils identifiés initialement pour la réalisation de l'intervention, 42 ont été supprimés, associés davantage aux profils islamistes qu'à ceux d'extrême droite.

Une situation similaire a été rencontrée par le programme Redirect de Moonshot (Helmus et Klein, 2018). Redirect utilise «Google AdWord» pour identifier les personnes recherchant des contenus extrémistes violents sur la plateforme et les exposer, dans les résultats de recherche, à de la publicité sous forme de vidéos exposant des contre-narratifs. La campagne de Moonshot a été temporairement interrompue à plusieurs reprises lorsque Google a essayé de limiter la publicité sur les termes de recherche racistes. Helmus et Klein mentionnent également le fait que toute évaluation basée sur un programme de médias sociaux dépend de la nature autosélective des utilisatrices et utilisateurs qui choisissent le contenu à consommer et que les comparaisons entre les publics qui sont exposés à une campagne et ceux qui ne le sont pas peuvent ainsi être biaisées.

Ces difficultés démontrent l'importance de la coordination avec tous les acteurs et actrices clés, y compris les grandes compagnies technologiques, lors de la mise en place d'un programme de prévention, mais aussi lors de son évaluation.

e. Les limites externes

Cette dernière catégorie de limites concerne des facteurs externes au processus d'évaluation et du programme qui ont cependant eu un effet considérable sur l'interprétation des données récoltées. Nous en avons trouvé trois : les limites de temps, les limites contextuelles et de sécurité et les limites budgétaires.

Les limites de temps

Comme il a été évoqué précédemment, le temps est un facteur clé qui affecte les différentes dimensions et étapes d'une évaluation. C'est l'un des facteurs les plus mentionnés par les documents retenus (Azam et Bareeha, 2017 ; Bean et al., 2011 ; Boyle et al., 2016 ; Hiariej et al., 2017 ; Muncy et al., 2015 ; Schorn et al., 2010). De prime abord, il s'agit du temps alloué à l'évaluation qui est souvent, d'après les évaluatrices et évaluateurs, très court pour collecter suffisamment d'information ou se rendre sur tous les sites à évaluer. Hiariej et ses collaborateurs illustrent plus clairement la pression du temps sur les personnes évaluatrices lorsqu'elles ont collecté des données avec un sondage papier sur cinq sites différents en Indonésie, en plus de la révision de la documentation du programme, des entretiens et des groupes de discussion :

Le sondage ne peut être mené simultanément, car l'évaluation ne peut mobiliser qu'un nombre limité de chercheurs et d'assistants locaux. Le temps alloué pour le sondage était d'environ 5-6 jours pour la collecte des données dans chaque ville. Si l'on ajoute à cela l'intervalle de 1 à 2 jours entre chaque sondage, les jours fériés scolaires et nationaux, ainsi que d'autres retards techniques sur le terrain, il a fallu près de 3 mois rien que pour collecter tous les questionnaires distribués. (2017, p. 17)

Cette courte période a des implications évidentes pour les évaluations. En particulier, seuls les effets immédiats peuvent être évalués alors que l'évaluation néglige les effets à plus long terme qui nécessitent des modèles longitudinaux.

Le temps alloué aux évaluations constitue seulement un des aspects de la dimension temporelle. Dans la partie précédente, nous avons aussi évoqué la période du déroulement du programme. Cependant, le *timing* entre la période d'évaluation et la disponibilité de certains acteurs et actrices sur le terrain est aussi important. Des situations anodines, comme la période des vacances scolaires ou les périodes de congé, peuvent avoir un impact sur le bon déroulement de l'évaluation (Hiariej et al., 2017 ; Nicolls et Hassan, 2014). Dans d'autres cas, c'est la disponibilité des personnes à interviewer, notamment des membres du gouvernement ou le temps limité des personnes qui ont mis en place l'intervention

(D. Parker et Lindekilde, 2020). L'évaluation est souvent, il faut l'avouer, une tâche connexe pour les actrices et acteurs de terrain sauf pour les personnes évaluatrices elles-mêmes.

Finalement, le temps est un facteur à considérer lorsqu'il s'agit d'une évaluation qui fait appel aux souvenirs des personnes pour reconstruire l'historique d'un programme (*biais de mémoire*) (Hiariej et al., 2017 ; Ipp et al., 2014 ; Khalil et al., 2019 ; Khalil et Ipp, 2016 ; Kollmorgen et Barry, 2017 ; Tines et al., 2017 ; Vittum et al., 2016). C'est le cas, notamment, des évaluations dont le design a considéré une étude de référence initiale pour comparer les résultats avant l'intervention qui n'existait pas et qui ont dû les reconstruire à partir de l'information fournie par les personnes intervenantes (Ipp et al., 2014). Ce sont aussi les cas où les évaluations ont été mises en place plusieurs mois après la fin du programme (Khalil et Ipp, 2016 ; Kollmorgen et Barry, 2017) ou lorsque les évaluatrices et évaluateurs ont dû poser des questions sur la biographie des personnes participantes ou sur le statut de la personne avant le démarrage du programme (Khalil et al., 2019 ; Tines et al., 2017). Khalil et ses collègues (2019) ont, par exemple, recueilli des informations auprès d'individus qui se sont désengagés d'Al-Shabaab plusieurs années auparavant pour évaluer un programme de désengagement.

Les limites politiques et sécuritaires

Bien que l'extrémisme violent soit un phénomène politique, il est rare que la dimension politique soit abordée comme un facteur dans l'évaluation des programmes. Le seul auteur qui mentionne cette dimension est Sarota (2017) qui considère que les biais politiques ont contribué à ce que certains répondants et répondantes ne donnent pas leur avis de manière totalement ouverte. En revanche, la sécurité a été souvent suggérée comme un facteur clé, notamment dans le cas des pays en conflit ouvert ou en situation de postconflit. Il s'agit, notamment, de certains sites qui ont été inaccessibles à cause de cette situation (Bean et al., 2011 ; Khalil et Ipp, 2016 ; Pipe et al., 2016 ; Tesfaye et al., 2018), voire de chercheurs et chercheuses qui n'ont pas pu réaliser le travail directement sur le terrain « en raison des mauvaises conditions de sécurité et des défis logistiques dans certaines parties de la Somalie, l'équipe d'évaluation pourrait ne pas être en mesure de mener directement le travail sur le terrain » (Swedberg et Reisman, 2013, p. 14). Finalement, le contexte d'insécurité peut également rendre les répondantes et répondants plus méfiants à l'idée de répondre aux questions de personnes inconnues (Basse, 2018).

Les limites budgétaires

La dernière difficulté énoncée par certains auteurs et auteures a été la limite en ce qui concerne le budget. Malgré le fait que le budget est souvent associé à une

contrainte pour la recherche, il n'a été mentionné que par deux études (Bean et al., 2011; Mansour, 2017). Dans les deux cas, l'information est succincte et fait référence à l'impact que le montant alloué à l'évaluation a particulièrement eu sur la portée de la cueillette de données, c'est-à-dire sur le nombre d'entrevues réalisées ou le nombre de sites visités. D'après ces chercheurs et chercheuses, cette limitation a eu pour effet que les évaluations n'ont pas pleinement pu évaluer l'impact plus large du programme sur les bénéficiaires.

3.5 QUELLE EST LA QUALITÉ DES MÉTHODES UTILISÉES ?

L'un des enjeux du domaine mentionnés dans la section 1 de ce rapport a été la qualité des méthodes utilisées dans le domaine général de « security studies » et, en particulier, dans l'évaluation de programmes de prévention de l'extrémisme violent. Les analyses initiales de Silke (2001, 2006) dans le domaine général du contreterrorisme montraient, par exemple, que les études utilisaient, notamment, des données secondaires et qu'il existait un manque d'analyses quantitatives. Dans le cas spécifique de l'évaluation des programmes de prévention de l'extrémisme violent, une grande partie des recensions passées n'ont pas réalisé ce type d'analyse (Feddes et Gallucci, 2015; Gielen, 2017; Mastroe et Szmania, 2016), se sont basées sur une échelle globale qui ne rend pas nécessairement compte des particularités des différents designs présents dans le domaine (Bellasio et al., 2018; Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021) ou n'ont pris en considération que des études quantitatives (Carthy et al., 2020; Lum et al., 2006). Il en résulte un biais de design dans les analyses (quantitatif-expérimental) basé sur un type particulier d'évaluation (impact) et qui exclut en principe toute analyse de la qualité des études qualitatives et de processus. L'équipe de Bellasio (2018) a récemment utilisé, par exemple, la « Maryland Scientific Methods Scale » qui détermine la qualité en fonction du type de design et qui établit comme norme que les designs expérimentaux présentent une meilleure évaluation que les autres types de design, ce qui, comme nous le verrons plus tard, est loin d'être une réalité dans ce domaine d'étude.

Les 200 études et plus qui ont analysé des données primaires que nous avons identifiées sont la preuve que le domaine de l'évaluation de programmes de prévention de l'extrémisme violent a énormément évolué ainsi que sa qualité globale, du moins, en ce qui concerne le recours à des données primaires. Cependant, il est encore tôt pour conclure sur la qualité des résultats de ces études. Pour ce faire, nous avons besoin de pousser l'analyse encore plus loin et nous interroger en détail sur la qualité des méthodologies utilisées, qui est l'objectif de cette section.

Le contexte d'insécurité dans certains pays a rendu l'accès au terrain difficile, voire impossible dans certains cas.

Comme mentionné dans la méthode, nous avons utilisé le « Mixed Methods Appraisal Tool (MMAT) » (Hong et al., 2018; Hong et Pluye, 2019) comme outil d'évaluation. Cet outil permet d'évaluer différents designs méthodologiques, y compris le degré d'intégration des méthodologies qualitatives et quantitatives dans le cas des études mixtes. Ces designs sont des études expérimentales, quasi expérimentales, quantitatives descriptives, qualitatives et mixtes⁴⁴. Il est composé ainsi de cinq échelles qui se déclinent à leur tour en cinq questions dont la réponse a été codée comme 1 (présence de l'indicateur), 0 (absence de l'indicateur) ou « ce n'est pas possible de déterminer »⁴⁵. Pour analyser les résultats obtenus par étude, nous avons agrégé les cas où l'indicateur était présent⁴⁶. De cette manière les résultats comprennent une plage allant de 0 à 5. Étant donné que chaque étude peut potentiellement utiliser les cinq designs méthodologiques et que le nombre des designs utilisés peut varier d'une étude à une autre, nous n'avons pas analysé chaque étude de façon globale et individuelle, mais plutôt les sections qui correspondent à chaque type de design. Il en résulte que certaines études seront analysées plus d'une fois selon le nombre de designs utilisés.

3.5.1 Quelle est la qualité des études qualitatives ?

D'après le manuel du MMAT, les études qualitatives permettent « d'explorer et de comprendre le sens que les individus ou les groupes attribuent à un problème social ou humain » (Creswell, 2013b, cité dans Hong et al., 2018, p. 3). Nous avons identifié 188 études qui ont utilisé, au moins en partie, des méthodes qualitatives (Encadré 6).

⁴⁴ Une description de chaque type de design sera apportée dans l'introduction de chaque section.

⁴⁵ Cette alternative a été codée lorsque l'information n'a pas été présente dans l'étude (par exemple, dans le cas des études dont la section méthodologique n'a pas été présente ou a été très limitée) ou lorsque la nature de l'étude empêchait de déterminer la valeur de l'indicateur.

⁴⁶ Il n'est pas recommandé d'utiliser l'addition de scores comme critère d'inclusion dans le contexte d'une revue systématique qui s'intéresse aux effets des programmes évalués. Nous avons utilisé ce score non pas comme un critère d'inclusion, mais comme une variable à décrire dans le cadre de cette étude.

Encadré 6. Indicateurs des études qualitatives utilisés dans le MMAT

1. L'approche qualitative est-elle appropriée pour répondre à la question de recherche ?
2. Les méthodes de collecte de données qualitatives sont-elles adéquates pour répondre à la question de recherche ?
3. Les conclusions sont-elles correctement tirées des données ?
4. L'interprétation des résultats est-elle suffisamment corroborée par les données ?
5. Y a-t-il une cohérence entre les sources de données qualitatives, la collecte, l'analyse et l'interprétation ?

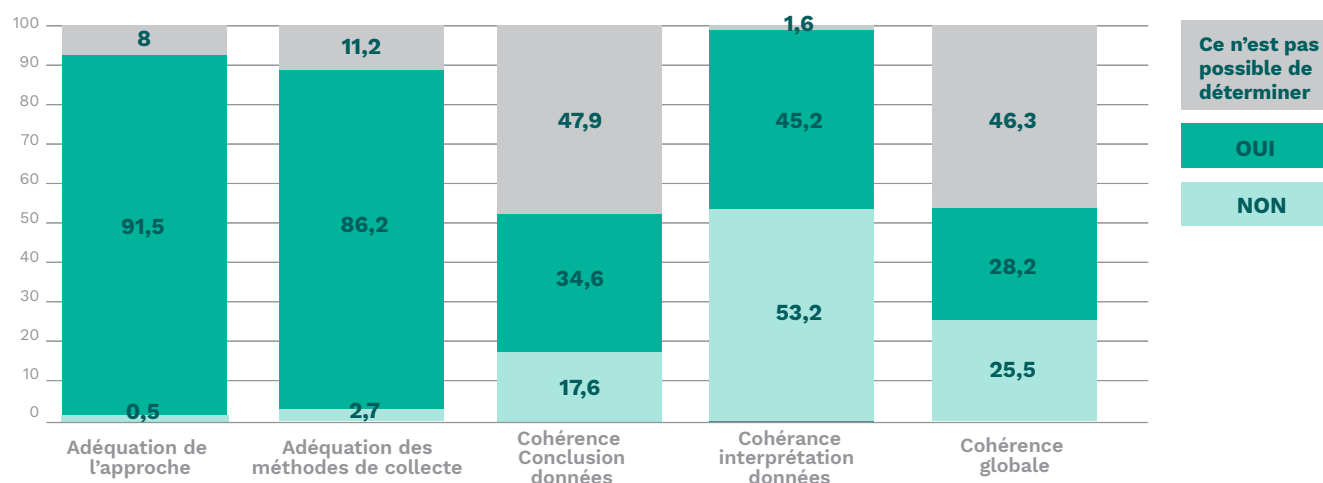
La qualité moyenne des études (2,86) est modérée, en étant légèrement au-dessus du milieu de l'échelle (2,5), la plupart des études (59 %) se concentrant sur une qualification de 2 ou 3. En revanche, 1 étude sur 4 (28,2 %) a été évaluée très positivement (4 ou 5, selon l'échelle). Si l'on regarde la Figure 17, on observe que le problème dans ce type d'étude réside davantage dans l'interprétation des données que sur l'adéquation de l'approche ou la collecte de données, ces derniers indicateurs étant évalués de façon très positive. Ce qui pose notamment un problème, c'est de savoir si l'interprétation des résultats a été suffisamment corroborée par les données. En revanche, dans le cas de l'association entre les données et la conclusion et la cohérence de l'ensemble de l'approche qualitative, le problème est plutôt associé à l'impossibilité de déterminer la présence ou l'absence de ces indicateurs. Ce fut le cas dans environ la moitié des études retenues. Dans le cas des études qualitatives, ces

trois derniers indicateurs peuvent être partiellement ou totalement corroborés par le biais de citations directes des personnes interviewées. Il peut donc être question dans ce contexte d'un problème d'espace. Les études qualitatives nécessitent généralement plus d'espace pour développer des idées et inclure des citations directes, ce qui n'est pas forcément le cas des études quantitatives. Les revues scientifiques limitent en revanche le nombre de mots de chaque publication, ce qui peut avoir un impact direct sur le nombre de détails à apporter. Cette interprétation est corroborée par la différence entre les études issues de la littérature académique et celles issues de la littérature grise, l'évaluation de cette dernière étant plus positive. Dans ce dernier cas, les limites d'espace ne posent pas un problème et les résultats peuvent inclure tous les détails nécessaires pour justifier l'interprétation des données.

Tableau 26. Qualité méthodologique des études qualitatives

Variable	Alternative	Nombre d'études	Moyenne	Écart type
Total		188	2,86	1,47
Type de publication	Littérature académique	67	2,69	1,62
	Littérature grise	121	2,95	1,38
Continent	Afrique	43	3,00	1,29
	Amérique du Nord	15	3,13	1,68
	Asie	37	2,11	1,49
	Europe	86	3,03	1,44
	N/A	2	2,00	1,41
	Océanie	5	3,60	1,34
Niveau de prévention	Primaire	38	2,63	1,63
	Primaire ciblée	75	3,12	1,50
	Secondaire	54	3,00	1,29
	Tertiaire	44	2,68	1,38
	Générale	21	2,76	1,58
Type d'extrémisme	Extrême droite	14	2,50	1,16
	Islamisme	77	3,04	1,45
	Non spécifique	109	2,80	1,48
Objectif de l'évaluation	Impact	130	2,92	1,39
	Processus	109	2,95	1,47
	D'extrants	24	3,25	1,45

Figure 17. Qualité méthodologique des études qualitatives par indicateur (%)



Par continent, l'Amérique du Nord est la région où les résultats sont les plus positifs (3,13), suivie par l'Europe (3,03) et l'Afrique (3). En revanche, l'Asie montre des résultats plus décevants quant à la qualité des méthodologies utilisées. Ces résultats sont positifs pour une région comme l'Afrique qui se montre à la hauteur d'autres régions qui ont une plus grande tradition en évaluation. Cependant, ils doivent être observés avec précaution, car ils peuvent également être biaisés. L'outil d'évaluation a été développé en Amérique du Nord en se basant sur des critères qui font consensus parmi les chercheuses et chercheurs dans cette région et en Europe, ce qui peut expliquer des résultats plus positifs des études dans ces deux continents. Une partie importante des auteures et auteurs qui ont évalué des programmes en Afrique, comme nous l'avons vu dans la section correspondante (3.2.2), provenaient également de l'Europe et, particulièrement, des États-Unis. Ils ont été formés dans cette tradition et ont utilisé ces critères pour la mise en place des évaluations. Cependant, bien que la méthode ait été transparente dans de nombreuses études, d'autres enjeux mettent un voile d'incertitude quant à la qualité des données cueillies et à leur interprétation, notamment dans le cas des études en Afrique. Cet outil ne considère pas, par exemple, d'autres critères de base qui, d'après notre point de vue, sont aussi fondamentaux pour évaluer la qualité d'une étude. C'est le cas, notamment, des évaluatrices et évaluateurs qui ne parlaient pas la langue du pays où ils faisaient l'évaluation et de l'accès à l'information, qui ont été évoqués dans la section précédente.

Quant aux autres variables présentées dans le Tableau 26, les études qui ont ciblé des programmes de prévention primaire ciblée et secondaire montrent en moyenne une évaluation de la qualité méthodologique plus positive ainsi que les études qui ont ciblé les programmes visant l'extrémisme islamiste et celles dont l'un des objectifs était l'évaluation des extrants. Ce dernier cas est plus étrange, car ce type d'évaluation est souvent très descriptif et quantitatif. Toutefois, les études qui visent à évaluer les extrants des programmes, entre autres objectifs, sont également celles qui correspondent aux rapports de littérature grise, de sorte que leurs évaluations positives peuvent être expliquées de la même manière.

Encadré 7. Deux exemples d'évaluations qualitatives

Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program (Savoia et al., 2016)

L'équipe de Savoia (2016) a évalué le cadre global du « *Greater Boston Countering Violent Extremism (CVE) Pilot Program* » qui visait à renforcer la résilience et la capacité de diverses communautés à empêcher que des individus, y compris des jeunes, ne soient inspirés et recrutés par des extrémistes violents. Le but de l'évaluation a été de collecter des données d'évaluation formative, au travers d'une perspective de santé publique, concernant les objectifs du programme et les recommandations sur la façon dont le programme devrait évoluer. Ils se sont basés sur trois objectifs spécifiques :

- Recueillir des avis sur les objectifs du programme et sur l'initiative globale.
- Identifier les recommandations pour la pratique.
- Élaborer un modèle logique pour l'évaluation des activités de prévention de la violence en fonction de la demande de subvention.

Par conséquent, ils n'ont pas évalué les activités mises en place, mais plutôt le cadre du projet.

Pour ce faire, ils ont interviewé des individus ayant une variété de perspectives et d'expériences liées au programme et qui ont été identifiés par l'intermédiaire de la technique de boule de neige. L'information fut analysée à travers un système de codage afin de répondre aux trois objectifs spécifiques. Toutes les étapes de design, de collecte et d'analyse ont été bien expliquées à l'exception des sous-dimensions abordées au sein de chaque objectif. Les interprétations de l'information ont toujours été illustrées par des citations directes des personnes interviewées.

« Prevent » in Southwark – 2009-2010 Evaluation Report (Rooke et Slater, 2010)

Un autre exemple est l'évaluation de la mise en œuvre d'un projet local dans le quartier de Southwark, à Londres, dans le contexte de la stratégie nationale du Royaume-Uni « Prevent ». Ce projet a donné la priorité à l'engagement des communautés musulmanes locales, à la compréhension de leurs besoins, à la création d'un réseau solide d'organisations musulmanes travaillant en partenariat avec les agences locales par le biais d'une approche de développement communautaire dialogique afin de prévenir l'extrémisme violent dans le quartier. L'évaluation a eu pour objectif d'évaluer la réalisation du projet au cours de l'année écoulée, de mettre en évidence les réussites et d'explorer les raisons de celles-ci et d'examiner le rapport qualité-prix par rapport à la réalisation du projet, parmi d'autres objectifs. Afin d'évaluer ce projet, les évaluatrices et évaluateurs ont mené des entrevues avec le personnel clé, ont réalisé deux études de cas sur une émission de radio destinée aux jeunes musulmans du sud de Londres et un documentaire télévisé sur l'islam à Southwar. D'après l'évaluation, ce projet a été une réussite, car il a pu adresser l'ensemble des priorités et recommandations à compléter. Parmi ces résultats, nous pouvons mentionner la réalisation d'un nombre important de projets axés sur la jeunesse, un développement intergénérationnel entre les jeunes et les personnes âgées, une création de réseaux entre les groupes musulmans de Southwark. L'évaluation, pour sa part, a réussi à rendre cohérente toute la démarche méthodologique et a eu du succès là où d'autres études qualitatives ont eu plus de problèmes : la relation entre les données recueillies et l'interprétation de ce matériel. En revanche, ils ne mentionnent pas les limites de l'évaluation.

3.5.2. Quelle est la qualité des études quantitatives descriptives ?

Le MMAT définit les études quantitatives descriptives comme les études qui sont « conçues uniquement pour décrire la distribution existante des variables sans trop se soucier des relations causales ou d'autres hypothèses » (Porta et al., 2014, p. 72, cités dans Hong et al., 2018). Nous avons identifié 60 études qui ont eu au moins une section quantitative descriptive (Tableau 27 et Figure 18).

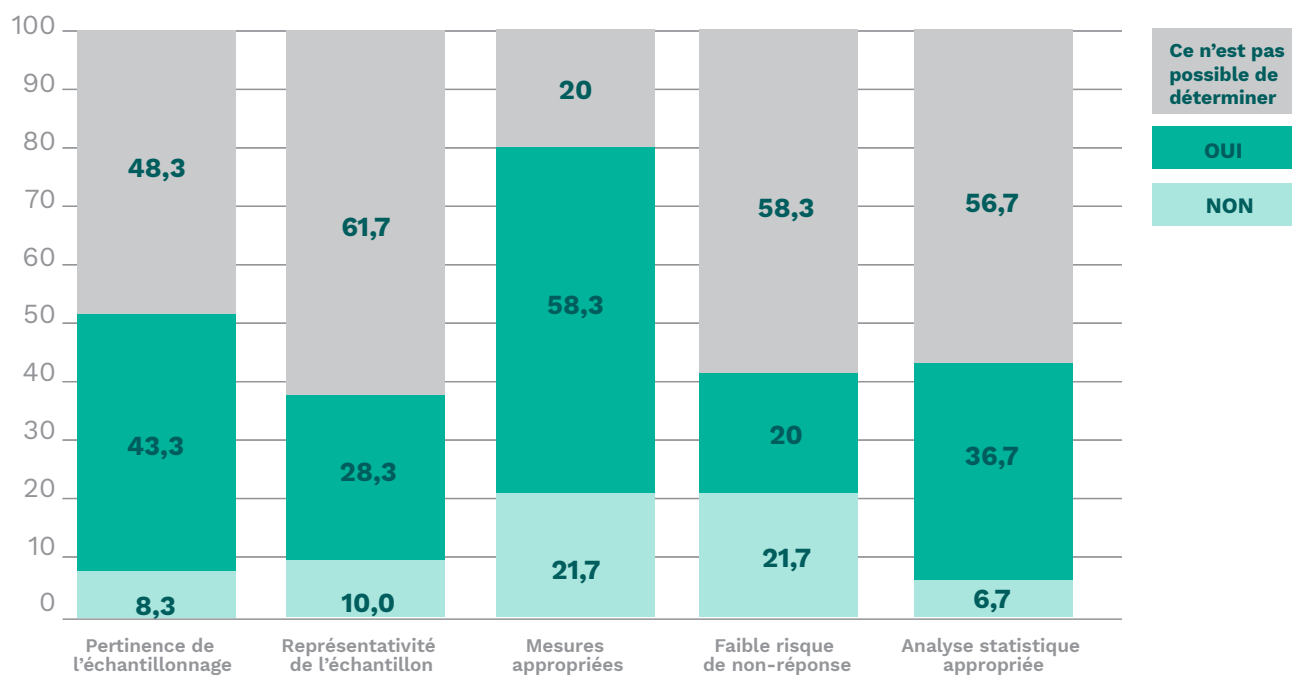
Encadré 8. Indicateurs des études quantitatives descriptives utilisés dans le MMAT

1. La stratégie d'échantillonnage est-elle pertinente pour répondre à la question de recherche ?
2. L'échantillon est-il représentatif de la population cible ?
3. Les mesures sont-elles appropriées ?
4. Le risque de biais de non-réponse est-il faible ?
5. L'analyse statistique est-elle appropriée pour répondre à la question de recherche ?

Tableau 27. Qualité méthodologique des études quantitatives descriptives

Variable	Alternative	Nombre d'études	Moyenne	Écart type
Total		60	1,87	1,44
Type de publication	Littérature académique	22	1,27	1,12
	Littérature grise	38	2,21	1,51
Continent	Afrique	12	2,42	1,56
	Amérique du Nord	8	1,00	0,53
	Asie	11	1,45	1,29
	Europe	19	1,84	1,61
	N/A	4	3,25	0,96
	Océanie	6	1,83	1,33
Niveau de prévention	Primaire	16	2,00	1,46
	Primaire ciblée	28	1,57	1,14
	Secondaire	23	2,22	1,48
	Tertiaire	12	2,33	1,78
	Générale	1	3,00	
Type d'extrémisme	Extrême droite	7	2,00	1,73
	Islamisme	28	1,93	1,30
	Non spécifique	30	1,83	1,53
Objectif de l'évaluation	Impact	49	1,90	1,39
	Processus	30	2,03	1,52
	D'extrants	11	1,91	1,30

Figure 18. Qualité méthodologique des études quantitatives descriptives par indicateur (%)



Les études qui ont au moins une section quantitative descriptive ont été en moyenne moins positivement évaluées (1,87) que les études qualitatives et en dessous du milieu de l'échelle (2,5), la grande majorité étant qualifiée entre 0 et 2 (66,7 %). À peine 15 % des études ont eu une qualification très positive (4 ou 5). Le principal problème de ces études est la transparence des procédures. Par exemple, dans 6 études sur 10, nous n'avons pas pu obtenir l'information nécessaire pour déterminer si l'échantillon a été représentatif, s'il existait un risque de non-réponse ou si les analyses étaient appropriées et dans 5 études sur 10, nous n'avons pas pu déterminer la pertinence de l'échantillonnage. Il s'agit, notamment, d'études qui ont présenté des résultats sans se soucier beaucoup de présenter également la façon dont ces résultats ont été obtenus. Ces résultats sont d'autant plus inquiétants que 84 % de ces études avaient parmi leurs objectifs globaux de mesurer l'impact du projet⁴⁷.

Dans les cas où nous avons pu déterminer la présence ou l'absence de l'indicateur, les résultats ne sont pas non plus encourageants. Le cas le plus positif est celui de la pertinence des mesures utilisées qui atteint 58,3 % de toutes les études. Il s'agit particulièrement de la validité et de la fiabilité de ces mesures et de savoir si les mesures ont été appropriées pour répondre à la question de recherche. C'est notamment ce dernier cas qui a pu être déterminé dans notre évaluation. Quant à l'échantillonnage, il a été pertinent dans 43,3 % des cas et il a été représentatif dans seulement 28,3 % des études. Ce dernier résultat doit toutefois être analysé avec précaution. Les critères du MMAT n'impliquent pas nécessairement un calcul de la représentativité de l'échantillon en relation à la population cible, mais de savoir s'il y a une correspondance entre les personnes

répondantes et la population cible ou s'il y a eu une description claire de la population cible et de l'échantillon, entre autres critères. Ce sont ainsi des critères larges qui ont été retrouvés dans seulement 28,3 % des études retenues. Dans la même veine, la moitié des études sur lesquelles nous avons pu déterminer le risque de non-réponse (41,7 %) présentaient un faible taux de réponse, ce qui limite évidemment la qualité des analyses réalisées même si, dans un tiers de celles-ci, l'analyse était pertinente pour répondre à la question de recherche.

Ces résultats sont clairement meilleurs dans le cas de la littérature grise, ce qui est en concordance avec les résultats des études ou sections qualitatives ainsi que dans le cas des études concernant l'Afrique et les programmes de prévention tertiaire et secondaire. Les résultats de l'Afrique sont aussi cohérents avec ceux des études qualitatives et les mêmes précautions doivent être prises lors de l'analyse de ces résultats. Les études africaines sont plus fortes dans le cas de la cohérence des mesures prises par rapport aux questions de recherche et beaucoup moins dans le cas de la représentativité de l'échantillon. Dans ce dernier cas, nous n'avons pas pu déterminer la présence ou l'absence de cet indicateur dans 7 études sur 10. En revanche, les études descriptives de l'Amérique du Nord ont les qualifications les plus basses de tous les continents. Dans 7 études sur 8, par exemple, nous n'avons pas pu déterminer le taux de réponse et, dans la seule qui reste, les taux étaient bas. Les évaluations de programmes de prévention tertiaire et secondaire présentent de meilleurs résultats que les programmes travaillant à d'autres niveaux de prévention dans le cas de la cohérence des mesures prises, ce qui suit la même tendance générale, mais aussi dans la pertinence de l'échantillon.

Encadré 9. Un exemple d'une évaluation quantitative descriptive

Une évaluation de « *Prevent* » dans le milieu éducatif (Joyce, 2018)

Joyce a exploré les perceptions et attitudes des enseignantes et enseignants envers la mise en œuvre du programme « *Prevent* » dans le milieu scolaire à West Yorkshire, en Angleterre. Il a utilisé un design mixte explicatif séquentiel qui incluait une partie descriptive quantitative et une autre qualitative. La population cible de la section quantitative était composée de personnes enseignantes de 10 écoles primaires et de 2 écoles secondaires, dont l'échantillon final fut de 38 enseignantes et enseignants qui ont répondu à un court sondage de 14 questions. Les données quantitatives ont été analysées à l'aide de statistiques descriptives. Cette étude, vue depuis l'angle du MMAT, présente toutes les caractéristiques d'une bonne étude quantitative-descriptive. Plus précisément, l'échantillonnage a été cohérent avec les questions de recherche, les mesures ont été appropriées et les analyses cohérentes avec l'approche proposée et, en général, la méthode a été transparente et bien expliquée. On peut lui reprocher le manque d'information quant aux caractéristiques de la population cible (combien de personnes enseignantes travaillaient, par exemple, au total dans ces écoles) afin de savoir avec plus de précision si son échantillon était représentatif de cette population.

Parmi les résultats obtenus, Joyce mentionne le besoin des enseignantes et enseignants que les éléments de base du programme « *Prevent* » soient plus clairs ainsi que leur besoin de soutien et de formation.

⁴⁷ Il convient de rappeler que, comme indiqué dans la section sur les types d'évaluation (section 3.3.1), l'objectif de l'évaluation d'impact a été déterminé par la déclaration explicite des auteures et auteurs dans les publications sélectionnées (objectifs, questions de recherche, intention déclarée, etc.) et non par une évaluation réalisée par les membres de cette étude.

3.5.3. Quelle est la qualité des études expérimentales ?

Les études expérimentales sont les études dans lesquelles les participantes et participants sont distribués de façon aléatoire dans des groupes d'intervention ou témoin, autrement dit, l'intervention est assignée par les chercheuses et chercheurs (Hong et al., 2018). Comme déjà susmentionné, nous avons identifié 6 études évaluatives qui peuvent être considérées comme expérimentales (Tableau 28).

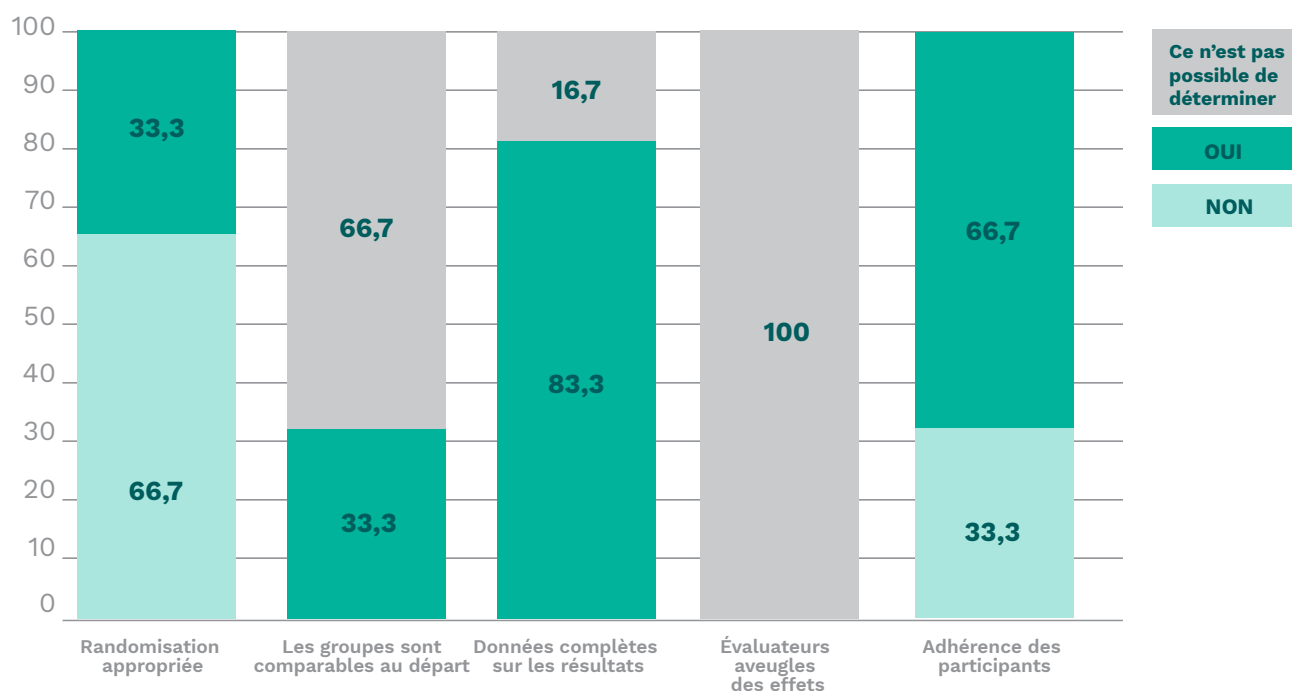
Encadré 10. Indicateurs des études expérimentales utilisés dans le MMAT

1. La randomisation est-elle effectuée de manière appropriée ?
2. Les groupes sont-ils comparables au départ ?
3. Les données sur les résultats sont-elles complètes ?
4. Les évaluatrices et évaluateurs des effets de l'intervention sont-ils aveugles quant à l'intervention fournie ?
5. Les personnes participantes ont-elles adhéré à l'intervention assignée ?

Tableau 28. Qualité méthodologique des études expérimentales

Variable	Alternative	Nombre d'études	Moyenne	Écart type
Total		6	2,17	1,47
Type de publication	Littérature académique	3	2,33	1,53
	Littérature grise	3	2,00	1,73
Continent	Afrique	2	1,00	0,00
	Amérique du Nord	0		
	Asie	1	1,00	
	Europe	2	3,00	1,41
	N/A	1	4,00	
	Océanie	0		
Niveau de prévention	Primaire	3	2,00	1,73
	Primaire ciblée	5	1,80	1,30
	Secondaire	0		
	Tertiaire	0		
	Générale	0		
Type d'extrémisme	Extrême droite	1	2,00	
	Islamisme	1	1,00	
	Non spécifique	4	2,50	1,73
Objectif de l'évaluation	Impact	6	2,17	1,47
	Processus	1	2,00	
	D'extrants			

Figure 19. Qualité méthodologique des études expérimentales par indicateur (%)



La présence d'études expérimentales est probablement l'un des grands avancements des dernières années dans le domaine. Sous certaines conditions, ce type de design peut être utilisé pour évaluer des programmes de prévention de l'extrémisme violent, particulièrement dans le cas de programmes de prévention primaire ou qui ne ciblent pas spécifiquement un type d'extrémisme. Toutefois, les résultats de l'évaluation de la qualité méthodologique doivent être analysés avec précaution (Figure 19). Le MMAT, bien que compréhensif, a été conçu pour évaluer n'importe quelle étude qui remplit certaines conditions dans le contexte de n'importe quel domaine. Certains indicateurs de l'évaluation d'études expérimentales s'appliquent ainsi moins bien que d'autres dans le contexte d'interventions sociales. C'est le cas, notamment, de l'aveuglement des personnes qui évaluent les effets de l'intervention. Dans un essai clinique randomisé en médecine, cet aveuglement implique que la personne qui évalue les effets d'un nouveau médicament ne connaît pas forcément si l'individu a reçu le placebo ou le médicament. Dans le cas des études qui évaluent les interventions sociales, la personne qui évalue les effets est souvent la personne qui participe à l'étude par l'intermédiaire d'un questionnaire. La personne sait évidemment si elle a participé ou pas à une intervention. L'aveuglement est ainsi souvent impossible⁴⁸.

Dans notre cas, dans aucune des six études expérimentales identifiées, nous n'avons pu déterminer un tel aveuglement. L'échelle a donc été modifiée pour considérer seulement

quatre des cinq critères. Sur cette base, le milieu de l'échelle change. Si pour les autres types de design, les études ont été cotées de 0 à 5 (moyenne=2,5), pour les études expérimentales, l'échelle de référence sera comprise entre 0 et 4, le point central ou la moyenne étant deux.

La moyenne des six études est de 2,17, légèrement au-dessus du milieu de l'échelle, dont quatre ont été qualifiées entre 1 et 2 et les deux restantes avec un score de 4. Les résultats sont ainsi modérés. Deux indicateurs ont eu une performance limitée : le processus d'assignation randomisée au groupe témoin et au groupe d'intervention et la comparabilité de ces groupes au départ. Ce sont deux éléments de base de ce type d'étude. Le MMAT a des critères plus exigeants quant à la randomisation : « Une simple déclaration telle que "nous avons procédé à une répartition aléatoire" ou "en utilisant un design aléatoire" est insuffisante pour juger si la randomisation a été effectuée de manière appropriée »⁴⁹ (Hong et al., 2018, p.4). Il faut donc avoir un plan ou une séquence prédéterminée bien expliquée. Dans les deux tiers de ces études, cette explication a été insuffisante. Dans le deuxième indicateur, la situation est semblable, cependant, dans ce cas, l'information quant à la comparabilité des groupes n'a pas pu être déterminée par notre équipe. Cette information est vitale pour déterminer si les changements observés sont dus à l'intervention ou aux caractéristiques des groupes étudiés.

⁴⁸ Une exception est l'utilisation d'activités de sensibilisation ou de formation comme « placebo » pour le groupe témoin.

⁴⁹ A simple statement such as 'we randomly allocated' or 'using a randomized design' is insufficient to judge if randomization was appropriately performed.

Les deux autres indicateurs ont obtenu des résultats plus positifs et concernent l'adhésion des participantes et participants à l'intervention et le degré dans lequel toutes les personnes participantes ont contribué à presque toutes les mesures (données complètes sur les résultats). Dans le premier cas, dans les deux tiers des études, l'adhérence n'a pas été un problème pour l'évaluation et, dans le deuxième cas, cette information n'a pas pu être déterminée dans seulement une étude. Le MMAT ne propose pas un seul standard ou seuil pour des données complètes acceptables sur les résultats, mais suggère d'utiliser un seul critère avec toutes les

études. Nous avons considéré le seuil le plus bas donné comme exemple (80 %).

Le faible nombre d'études nous empêche d'approfondir davantage sur une comparaison par différentes variables. Cependant, nous pouvons mentionner que, dans ce contexte, les études publiées dans des revues académiques et retenues dans cette revue systématique semblent être mieux qualifiées que celles issues de la littérature grise. Dans ce cas, l'évaluation des pairs semble être un facteur clé pour assurer une bonne qualité méthodologique.

Encadré 11. Trois évaluations expérimentales

Voices for Peace' Impact Evaluation of a Radio Drama to Counteract Violent Extremism in the Sahel Region in Burkina Faso (Bilali, 2019)

« Voices for Peace » est une intervention de 5 ans visant la réduction de la vulnérabilité à l'extrémisme violent dans la région du Sahel en Afrique de l'Ouest (Burkina Faso, Niger, Cameroun, Mali et Tchad). Les objectifs sont les suivants : 1) dénoncer l'extrémisme violent et réduire le soutien dont il bénéficie, 2) sensibiliser aux facteurs qui contribuent à l'extrémisme violent et au recrutement des jeunes, 3) accroître l'engagement des populations en lien avec le contre extrémisme et 4) encourager la gouvernance participative. Les auteures et auteurs évaluent donc par l'entremise de cette étude une partie de ce programme prenant la forme de divertissement éducatif par la radio, au Burkina Faso. Plus précisément, une émission de radio intitulée Wuro Potal se concentre sur l'extrémisme violent (en particulier la violence infligée à une communauté fictive par un groupe armé), la collaboration entre la population et les forces de sécurité/militaires, la gouvernance et la corruption, et la migration. Afin d'évaluer ses effets, 22 participantes et participants furent sélectionnés aléatoirement dans 132 villages afin d'écouter 52 épisodes éducatifs répartis sur 12 semaines. De plus, 66 de ces 132 villages furent assignés au programme tandis que les 66 autres ont formé le groupe témoin. Les résultats montrent que l'intervention a eu un très faible impact sur la réduction de la justification de la violence extrémiste, mais a augmenté les intentions comportementales de collaboration avec la police et les forces de sécurité, la sensibilisation à la gouvernance et à l'extrémisme violent et la perception de la capacité personnelle et collective à apporter des changements positifs et à améliorer les conditions de la communauté.

A former right-wing extremist in school-based prevention work: Research findings from Germany (Walsh et Gansewig, 2019)

Ce document résume les résultats d'une évaluation d'impact d'un programme de prévention de l'extrémisme violent (PVE) mené en milieu scolaire par un ancien extrémiste de droite. Plus précisément, le programme s'adressait aux élèves de huitième année ou plus et prenait la forme de quatre leçons (180 minutes) sur les thèmes de l'extrémisme de droite et de la criminalité. Ainsi, l'ancien extrémiste aborde des éléments théoriques ainsi que son expérience devant les jeunes, ce qui est par la suite suivi d'une discussion ouverte. L'objectif de cette mesure de prévention primaire était, entre autres, de prévenir l'extrémisme de droite et les comportements délinquants chez les jeunes. Afin d'évaluer le programme, un échantillon composé de 564 élèves provenant de 50 classes scolaires allemandes a été réparti également et aléatoirement entre un groupe de traitement et un groupe témoin. Les données récoltées par le biais de questionnaires et d'observations en classe ne suggèrent pas que le programme ait influencé les attitudes vis-à-vis l'extrémisme de droite et la délinquance. Toutefois, ces résultats ne sont pas considérés comme surprenants par les auteurs puisque, entre autres raisons, il ne fallait pas s'attendre à ce que les opinions ou le comportement des participantes et participants changent à la suite de la mise en œuvre d'une seule mesure de prévention de trois heures.

Preventing Extremism with Extremists: A Double-Edged Sword ? An Analysis of the Impact of Using Former Extremists in Danish Schools (D. Parker et L. Lindekilde, 2020)

À la demande des autorités danoises, dans cette étude, les auteurs évaluent l'efficacité d'une initiative financée par l'État. En effet, cette initiative vise la visite d'anciens extrémistes dans des écoles, des théâtres locaux et des centres jeunesse à travers le Danemark afin d'aborder leurs expériences et, ainsi, sensibiliser plusieurs jeunes âgés de 13 à 20 ans. Plus précisément, les ex-extrémistes doivent souligner les impacts négatifs de l'extrémisme violent sur leur vie, tout en décrivant comment ils ont d'abord été exposés aux idéologies extrémistes. Ainsi,

l'intervention tente d'aborder et de mettre en lumière le processus de radicalisation ainsi que ses effets néfastes, dans un but ultime de contrer les discours extrémistes et d'accroître la pensée critique des jeunes. Afin d'évaluer ce programme, 1931 jeunes personnes danoises furent sélectionnées afin de répondre à un questionnaire et furent également divisées en deux groupes distincts : 976 dans le groupe expérimental et 955 dans le groupe témoin. Les résultats indiquèrent, entre autres, que le projet s'est avéré efficace pour accroître la capacité des personnes participantes à reconnaître les idées extrémistes et les méthodes de recrutement. Cependant, les auteurs ont constaté une légère diminution quant à la tolérance politique pour les jeunes ayant participé au programme. En bref, l'évaluation apporte un certain soutien aux affirmations principalement théoriques de la littérature selon lesquelles les anciennes personnes extrémistes peuvent être des partenaires crédibles et efficaces pour la mise en œuvre de stratégies contre-terroristes. Cependant, les effets mitigés rappellent les risques de conséquences involontaires de telles initiatives, en particulier celles qui cherchent à influencer les attitudes.

3.5.4 Quelle est la qualité des études quasi expérimentales ?

Ces études sont définies comme toutes les études quantitatives estimant l'efficacité d'une intervention, qui n'utilisent pas la randomisation pour distribuer les participantes et participants dans les groupes témoin (Hong et al., 2018). Nous avons identifié 54 études qui remplissent ces conditions.

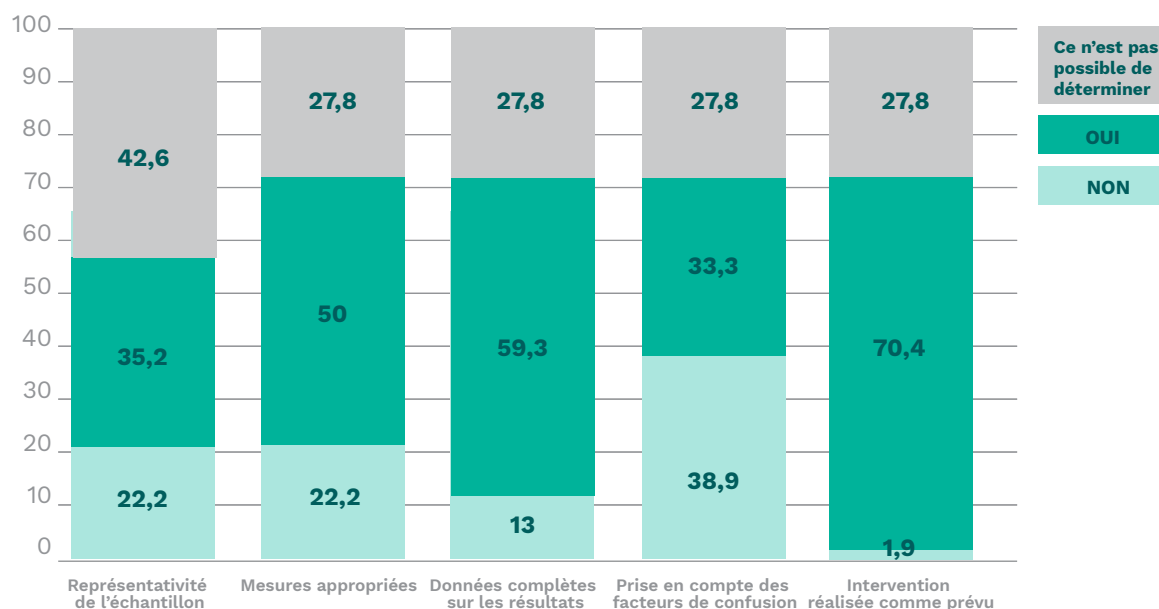
Encadré 12. Indicateurs des études quasi expérimentales utilisés dans le MMAT

1. Les personnes participantes sont-elles représentatives de la population cible ?
2. Les mesures sont-elles appropriées, tant en ce qui concerne le résultat que l'intervention ?
3. Les données sur les résultats sont-elles complètes ?
4. Les facteurs de confusion sont-ils pris en compte dans la conception et l'analyse ?
5. Pendant la période d'étude, l'intervention est-elle administrée comme prévu ?

Tableau 29. Qualité méthodologique des études quasi expérimentales

Variable	Alternative	Nombre d'études	Moyenne	Écart type
Total		54	2,48	1,69
Type de publication	Littérature académique	17	2,76	1,82
	Littérature grise	37	2,35	1,64
Continent	Afrique	16	2,88	1,54
	Amérique du Nord	3	4,00	0,00
	Asie	9	2,78	2,11
	Europe	25	1,96	1,62
	N/A			.
	Océanie	1	2,00	
Niveau de prévention	Primaire	11	2,36	1,80
	Primaire ciblée	22	2,86	1,81
	Secondaire	16	1,94	1,44
	Tertiaire	6	1,50	2,35
	Générale	8	2,63	1,19
Type d'extrémisme	Extrême droite	5	0,60	0,89
	Islamisme	13	3,46	1,61
	Non spécifique	37	2,16	1,55
Objectif de l'évaluation	Impact	54	2,48	1,69
	Processus	12	2,25	1,60
	D'extrants	5	2,00	0,00

Figure 20. Qualité méthodologique des études quasi expérimentales par indicateur (%)



La qualité moyenne de ces études quasi expérimentales se situe légèrement en dessous du milieu de l'échelle (2,48 sur 5) et y est relativement bien répartie : un tiers (27,8 %) a été évalué entre 0 et 1, un autre tiers (38,9 %), entre 2 et 3, et un autre tiers (33,4 %), entre 4 et 5 (Tableau 29 et Figure 20). Les problèmes se focalisent sur deux indicateurs : la représentativité de l'échantillon et la prise en compte de facteurs de confusion. Dans les deux cas, seul un tiers des études répond positivement à ces indicateurs. L'indicateur de représentativité de l'échantillon dans ces études, bien que légèrement plus strict que dans le cas des études quantitatives descriptives, est également évalué dans le MMAT de manière large et ne considère pas comme nécessaire le calcul de l'échantillonnage. Il suffit de décrire clairement la population cible et l'échantillon, y compris les critères d'inclusion et d'exclusion, et d'essayer au moins d'obtenir un échantillon représentatif. Cependant, nous n'avons pas été en mesure de déterminer la présence ou l'absence de l'indicateur dans 4 des 10 des études, ce qui constitue un problème grave de transparence méthodologique de ces études. Ce problème de transparence ne s'expliquerait pas par un souci d'espace, comme dans le cas des études qualitatives, puisque plus des deux tiers consistent en des rapports de littérature grise qui, en moyenne, ont été moins bien évalués que les études académiques. Il s'explique davantage par le fait que dans plus d'un tiers de ces études, l'échantillon n'était pas du tout décrit. Dans le cas du deuxième indicateur, il s'agit moins d'un problème de fond que d'un problème d'analyse et d'interprétation : près de 4 études sur 10 n'ont pas pris en compte des facteurs de confusion et, dans 3 études sur 10, nous n'avons pas pu déterminer cette information. Ces facteurs prédisent à la fois les résultats de la variable dépendante (les effets de l'intervention) et de la variable indépendante (l'intervention) et ne pas les considérer peut amener à une interprétation erronée de la causalité

entre l'intervention et ses effets. Bien que les autres indicateurs aient été plus présents dans les publications, les résultats ne sont pas excellents. Dans un tiers des études, par exemple, les trois indicateurs restants n'ont pas pu être déterminés. Seul l'indicateur évaluant si l'intervention a été réalisée comme prévu initialement montre des résultats positifs (70,4 %). En revanche, dans près des deux tiers des études (59,3 %), nous avons pu déterminer que la plupart des participantes et participants sectionnés ont contribué à la plupart des résultats obtenus (« données complètes sur les résultats ») et, dans la moitié des études (50 %), nous avons pu déterminer que les mesures ont été appropriées et validées pour répondre aux questions de recherche.

Ces résultats varient selon d'autres variables considérées. Les trois études quasi expérimentales de l'Amérique du Nord ont été qualifiées très positivement (4) ainsi que les treize études qui ont évalué des programmes de prévention de l'islamisme radical (3,46). En revanche, les cinq études qui ont évalué des programmes de prévention de l'extrême droite ont eu une très mauvaise qualification (0,6), notamment en raison des informations manquantes qui nous ont empêchés de déterminer la présence de l'indicateur. C'est le cas, comme nous l'avons vu, d'une grande partie des études qui ont eu une mauvaise qualification. Les six études qui ont évalué des programmes de prévention tertiaire montrent également une faible performance avec cet outil qui peut s'expliquer par les problèmes déjà mentionnés dans la section 1 de ce rapport, notamment le faible nombre de cas, l'absence de groupes témoins et le type d'indicateur utilisé

Encadré 13. Deux exemples d'évaluations quasi expérimentales

Preventing Violent Extremism through Value Complexity: Being Muslim Being British (Liht et Savage, 2013)

Liht et Savage (2013) ont développé et évalué le programme « *Being Muslim Being British* » qui a été conçu pour prévenir l'extrémisme violent chez les jeunes personnes musulmanes britanniques à travers le développement de la capacité des individus à comprendre les points de vue et valeurs des autres de façon plus complexe (complexité intégrative). Cette intervention a été évaluée auprès de 81 jeunes musulmanes et musulmans, à travers sept groupes pilotes au Royaume-Uni, avant et après leur participation à l'intervention. L'évaluation a été basée sur deux hypothèses : à la suite de l'intervention, les personnes participantes a) penseront de manière plus complexe et b) se montreront plus attentives à un plus grand nombre de valeurs (pluralisme des valeurs) lorsqu'elles travailleront sur des questions sociales caractérisées par des valeurs conflictuelles. Les indicateurs principaux découlent de ces hypothèses : la complexité intégrative et le style de résolution de conflits. Ces deux hypothèses ont été testées sur deux séries de données verbales qui ont été quantifiées : par le codage des réponses écrites à six dilemmes moraux auxquels les participantes et participants ont été exposés avant et après l'intervention et par le codage des discussions de groupe qui ont eu lieu pendant les activités de groupe au début et à la fin de l'activité. Ces informations ont été codées par deux personnes codeuses formées et aveugles aux conditions de préintervention et de post-intervention. Ce codage a été validé par l'intermédiaire du calcul de l'accord interjuge (Kappa de Cohen) entre les deux personnes codeuses. Les résultats de cette évaluation ont montré que la complexité intégrative des participantes et participants a augmenté de manière significative après l'intervention et que leur style de résolution des conflits est devenu plus collaboratif et engagé.

Mindanao Youth for Development (MYDev) Program (Education Development Center (EDC) et USAID, 2019)

Le projet MYDev, basé aux Philippines, était originalement un programme d'employabilité qui offrait une formation expérientielle et un soutien post-formation pour améliorer les compétences de vie, augmenter l'engagement civique et accroître l'employabilité de jeunes personnes vulnérables non scolarisées dans des zones touchées par le conflit armé dans le pays. Une année de prolongation lui a été accordée afin d'inclure un quatrième objectif relatif aux changements dans la perception des jeunes quant à la violence et l'extrémisme violent. L'Education Development Center (EDC) a été engagé par USAID afin de conduire une évaluation d'impact quasi expérimentale pour mieux comprendre la contribution de MYDev à ces objectifs. L'EDC a conduit cette évaluation sur deux cohortes de jeunes : la première concernait l'évaluation des trois premiers objectifs et la deuxième incluait des jeunes de l'année de prolongation. Dans les deux cas, il y a eu des prises de mesures avant et après l'intervention et seulement dans le cas de la première cohorte, l'équipe d'évaluation a eu recours à un groupe témoin. Dans le cas de la deuxième cohorte, les personnes évaluatrices ont conduit une technique d'échantillonnage aléatoire stratifié proportionnel pour choisir 789 jeunes du groupe d'intervention. Les données ont été recueillies au début des programmes de formation et 4 à 6 mois après la fin de ceux-ci. Pour la cueillette de données, ils ont utilisé deux outils validés et adaptés au contexte philippin (le Youth Employment Survey et le Youth Perceptions Survey) qui, ensemble, mesuraient huit indicateurs. Le premier outil a mesuré la préparation au travail, le leadership, les compétences de vie et les perceptions concernant le genre et, le deuxième outil, les perceptions du gouvernement, les perceptions de la communauté, les perceptions de la violence et les compétences en matière de résilience. Parmi les résultats obtenus, les jeunes qui ont participé au programme ont montré des changements positifs dans leur perception de la violence, y compris de l'extrémisme violent ainsi qu'une amélioration de leurs capacités de résilience.

3.5.5. Quelle est la qualité des études mixtes ?

Comme mentionné dans la section consacrée aux approches méthodologiques globales (page 49), la recherche par méthodes mixtes consiste à combiner des méthodes qualitatives et quantitatives. Dans le cas du MMAT, la qualité évaluée est associée, notamment, à l'intégration de deux méthodologies. Nous avons identifié 90 études qui ont utilisé une méthodologie mixte. Il faut noter que chaque section quantitative (descriptive, expérimentale ou quasi expérimentale) et qualitative de ces études a été déjà évaluée dans les sections précédentes de cette partie.

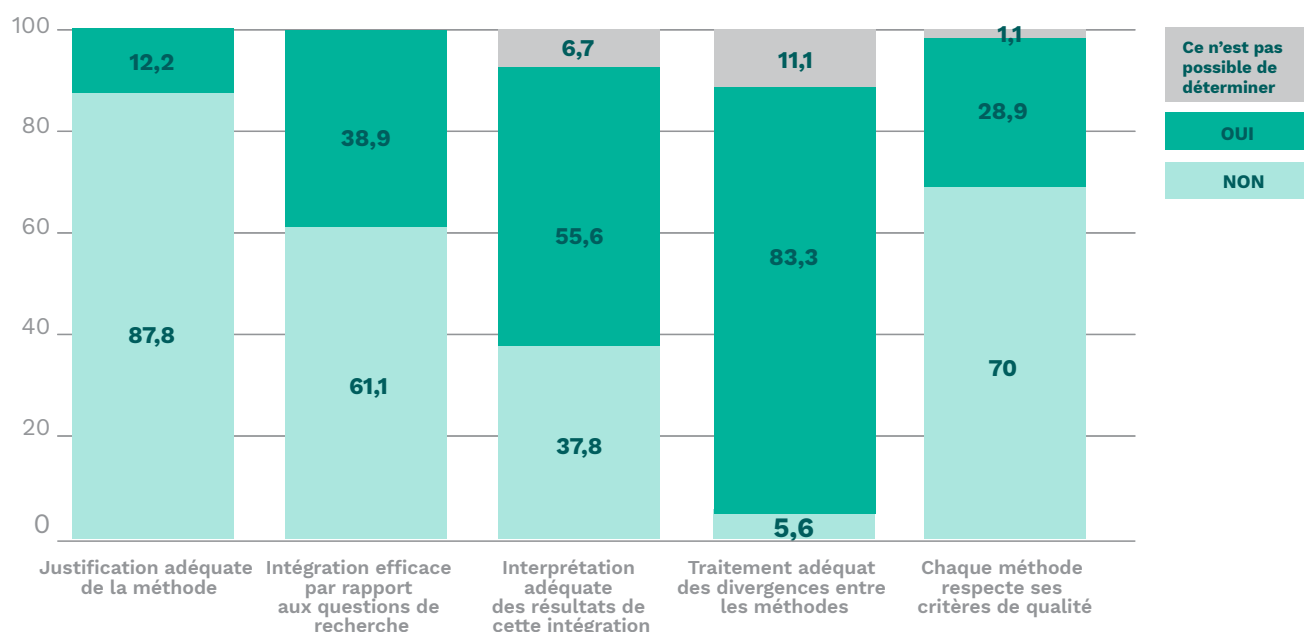
Encadré 14. Indicateurs des études mixtes utilisés dans le MMAT

1. Existe-t-il une justification adéquate de l'utilisation d'un modèle de méthodes mixtes pour répondre à la question de recherche ?
2. Les différentes composantes de l'étude sont-elles intégrées efficacement pour répondre à la question de recherche ?
3. Les résultats de l'intégration des composantes qualitatives et quantitatives sont-ils interprétés de manière adéquate ?
4. Les divergences et les incohérences entre les résultats quantitatifs et qualitatifs sont-elles traitées de manière adéquate ?
5. Les différentes composantes de l'étude respectent-elles les critères de qualité de chaque tradition des méthodes impliquées ?

Tableau 30. Qualité méthodologique des études mixtes

Variable	Alternative	Nombre d'études	Moyenne	Écart type
Total		90	2,19	1,47
Type de publication	Littérature académique	20	1,90	1,48
	Littérature grise	70	2,27	1,46
Continent	Afrique	25	2,52	1,23
	Amérique du Nord	8	1,63	1,06
	Asie	16	1,63	1,50
	Europe	34	2,38	1,61
	N/A	2	2,50	2,12
	Océanie	5	1,80	1,64
Niveau de prévention	Primaire	21	2,14	1,62
	Primaire ciblée	41	2,27	1,41
	Secondaire	32	2,28	1,46
	Tertiaire	16	2,25	1,88
	Générale	6	2,50	0,84
Type d'extrémisme	Extrême droite	7	1,86	1,21
	Islamisme	35	2,34	1,63
	Non spécifique	54	2,06	1,31
Objectif de l'évaluation	Impact	80	2,23	1,41
	Processus	42	2,43	1,53
	D'extrants	14	2,36	1,34

Figure 21. Qualité méthodologique des études mixtes par indicateur (%)



La qualité moyenne des études ayant utilisé des méthodes mixtes est en dessous du milieu de l'échelle (2,19 sur 5) dont presque la moitié (42,2 %) se concentre sur une qualification de 0 et 1 et seul un cinquième des études a été évalué de façon très positive (4 ou 5, selon l'échelle). Les indicateurs du MMAT sur ce type d'étude peuvent être résumés en trois catégories : justification, intégration et respect des critères propres de chaque méthode. Dans tous les cas, le problème de transparence méthodologique qui a été une constante dans les autres échelles est pratiquement inexistant ici : nous avons pu déterminer la présence ou l'absence de l'indicateur dans presque tous les cas. **La justification de la méthode est probablement le problème principal de ce type d'études : la plupart des études se contentent de mentionner l'utilisation d'une méthodologie combinée, mais justifient rarement le choix de ce type d'approche ou son avantage par rapport aux autres designs.** Les études mixtes sont devenues de plus en plus utilisées en sciences sociales et commencent à se considérer comme un modèle nécessaire pour répondre à des questions complexes. Cette normalisation peut expliquer ce manque de justification.

Le respect des critères propres à chaque méthode est la conséquence directe de l'évaluation précédente de chaque composante qualitative et quantitative (descriptive, expérimentale ou quasi expérimentale) de l'étude dans les sections correspondantes du MMAT. Dans le cas du MMAT, « la qualité globale d'une étude par méthodes mixtes ne peut pas dépasser la qualité de sa composante la plus faible » (Hong et al., 2018, p. 8). Dans cet indicateur, la qualité des deux composantes doit être élevée pour que l'étude mixte soit considérée comme

étant de bonne qualité. Dans notre cas, ce critère n'est rempli que dans 3 études sur 10.

En ce qui concerne l'évaluation de l'intégration de deux méthodes, les résultats sont hétérogènes. Le résultat le plus positif concerne, notamment, le traitement des divergences entre les composantes. Pour le MMAT, il y a deux options pour indiquer la présence de cet indicateur : il n'y a pas de divergences et lorsqu'ils existent, il ne faut pas seulement les mentionner, mais aussi les expliquer. Dans 8 études sur 10, nous avons constaté que ce critère était présent. Dans le cas de l'intégration globale des composantes dans l'interprétation des résultats, les conclusions sont modérées : nous avons trouvé la présence de cet indicateur dans seulement la moitié des études. Celui-ci est aussi associé à la justification du recours aux méthodes mixtes parce qu'il permet de « démontrer la valeur ajoutée de la conduite d'une étude à méthodes mixtes plutôt que de deux études distinctes » (Hong et al., 2018, p. 7). Le dernier indicateur associé à l'intégration de deux méthodes est de savoir si cette intégration répond aux questions de recherche. Les résultats dans ce cas sont moins positifs, ce critère étant présent dans seulement 38,9 % des études. Ce critère est le plus global de ceux qui ont été analysés au cours de ce paragraphe. Il nous montre la capacité des études à tracer un portrait global à partir des résultats des méthodes utilisées. Vues de cette manière, les études évaluatives mixtes nécessitent encore un peu d'amélioration afin de montrer tout le potentiel de la combinaison de méthodes.

Les résultats par des variables spécifiques varient également dans cette sous-échelle. Comme cela a été la tendance dans l'évaluation des études qualitatives

et quantitatives-descriptives, la littérature grise est mieux évaluée que la littérature académique dans cette échelle. C'est la situation contraire dans les cas des designs quantitatifs complexes tels que les designs expérimentaux et quasi expérimentaux.

Par continent, l'Afrique et l'Europe montrent des résultats plus positifs, mais les études asiatiques et nord-américaines ont été moins bien qualifiées.

Le type de prévention des programmes évalués ne semble pas être un facteur clé, les études qui ont évalué des programmes de prévention tertiaire et générale étant légèrement moins bien qualifiées. Cette tendance globale, c'est-à-dire l'influence du niveau de prévention sur la qualité des méthodes, se répète de façon plus accentuée dans les autres échelles analysées, notamment dans le cas de programmes de prévention tertiaire. Nous avons déjà mentionné les raisons qui expliquent cette situation dans la section précédente.

Par type d'extrémisme, la tendance se répète également : les études qui évaluent les programmes de prévention de l'islam radical sont systématiquement mieux notées que les autres, notamment celles qui évaluent les programmes destinés à l'extrême droite. Cette situation peut s'expliquer peut-être par l'expérience cumulée des études qui ciblent les programmes de prévention de l'extrémisme violent dit « islamiste », qui sont beaucoup plus nombreuses que celles qui ciblent les programmes de prévention de l'extrême droite. En dépit de cette affirmation, il est clair que ces dernières nécessitent encore de parcourir un chemin plus long pour être d'utilité pour la politique publique. Finalement, par type d'objectif d'évaluation, les évaluations, notamment des études mixtes, ne semblent pas varier énormément. En revanche, si l'on considère le reste des échelles, la situation est relativement différente : les évaluations d'impact sont systématiquement mieux qualifiées dans les cas des designs quantitatifs plus complexes (expérimentaux et quasi expérimentaux), ce qui est cohérent avec la question de recherche à laquelle elles essayent de répondre.

Encadré 15. Deux exemples d'évaluations de design mixte

“If Youth Are Given the Chance” Effects of Education and Civic Engagement on Somali Youth Support of Political Violence (Tesfaye et al., 2018)

Le programme de prévention « Mercy Corps' Somali Youth Learners Initiative (SYLI) visait à améliorer l'accès des jeunes Somaliens aux opportunités d'éducation et d'engagement civique comme un moyen de réduire leur probabilité de soutenir ou de se joindre à des groupes armés. L'équipe de Tesfaye (2018) a procédé à une évaluation d'impact de ce programme en utilisant un design mixte quasi expérimental. En plus de cette relation causale, l'équipe d'évaluateurs et d'évaluatrices a aussi testé des variables modératrices entre l'intervention proposée et les effets souhaités : le potentiel d'être déçu par les perspectives de moyens de subsistance, l'isolement social, la perception quant à la capacité à provoquer un changement communautaire, la confiance dans les moyens de résolution non violents, la confiance dans le gouvernement fédéral et la confiance dans le gouvernement de l'État. Ils ont également contrôlé certaines variables : le sexe, les indices de richesse et de faim, le niveau de violence dans la communauté, le niveau d'urbanisation et la durée de mise en œuvre de l'intervention.

Le design mixte consistait notamment à une enquête par sondage et à des entrevues auprès d'informatrices et d'informateurs clés. Dans le cas de l'enquête par sondage, l'équipe d'évaluateurs et d'évaluatrices a formé 3 groupes de comparaison : un groupe témoin (283 personnes participantes) ; un groupe d'intervention qui a suivi seulement le programme de formation (215 personnes participantes) et un deuxième groupe d'intervention qui a suivi le programme de formation ainsi que celui d'engagement civil (722 personnes participantes). Au total, 1220 jeunes personnes scolarisées et non scolarisées âgées de 15 à 24 ans y ont participé. Par la suite, des entrevues avec des informatrices et informateurs clés ont aussi eu lieu. Il y a eu 40 jeunes âgés de 18 à 30 ans qui ont été interrogés. Un guide d'entrevue, incluant des questions semi-dirigées et des éléments de narration, a été utilisé pour cette phase de collecte de données. Les deux composantes de la méthode ont été désignées afin de s'intégrer depuis le début. Les thématiques identifiées dans les entretiens ont été comparées aux résultats quantitatifs obtenus afin de mettre en évidence les points d'accord et de désaccord entre les sources de données et d'identifier des explications alternatives.

En somme, les chercheuses et chercheurs ont pu constater, dans l'ensemble, qu'un meilleur accès à l'enseignement secondaire soutenu par SYLI, avec ou sans engagement civique, a contribué à réduire les niveaux de soutien à la violence politique chez les jeunes Somaliens. Toutefois, lorsqu'il a été question de vérifier ce qui conduit aux réductions observées du soutien à la violence politique, ce n'est que certaines hypothèses qui ont été confirmées. Enfin, les résultats obtenus soulèvent d'autres questions importantes pour les futures recherches : par exemple, l'impact de l'ouverture d'une école ou dans quelle mesure l'éducation est primordiale dans l'augmentation des possibilités d'emploi à long terme par rapport à la création d'une résistance immédiate à la propagande extrémiste.

The role of self-help efforts in the reintegration of 'politically motivated' former prisoners: implications from the Northern Irish experience (Dwyer, 2010 ; Dwyer et Maruna, 2011)

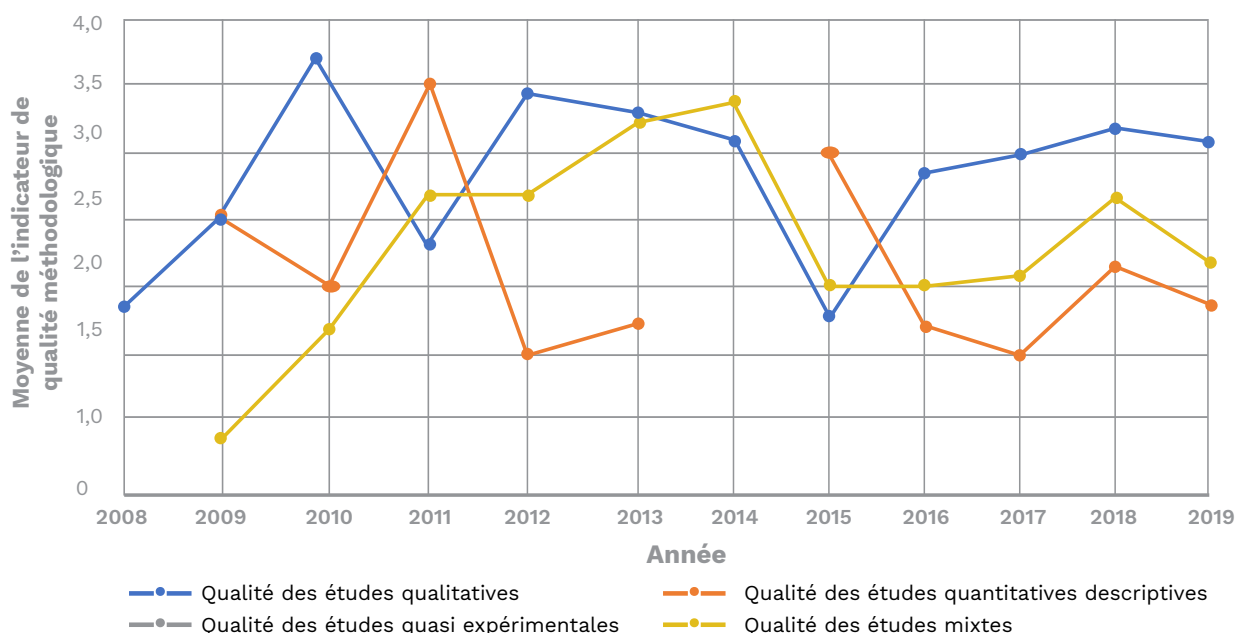
Cette étude a évalué les pratiques des organisations communautaires d'entraide dans le processus de réintégration des anciens prisonniers « à motivation politique » dans le conflit nord-irlandais. Pour ce faire, l'auteur a eu recours à une méthode mixte qui a combiné la réalisation de 35 entretiens avec d'anciennes personnes prisonnières, des membres d'organisations d'entraide, des représentantes et représentants du gouvernement et des membres d'organisations non gouvernementales avec un sondage auprès de 69 anciennes personnes prisonnières républicaines et loyalistes. Ces données ont également été complétées par des informations recueillies lors de nombreuses réunions et correspondances informelles avec des fonctionnaires, des universitaires et des membres du secteur communautaire/bénévole, notamment des organisations de défense des droits. L'approche mixte de Dwyer (2010) a été consécutive et exploratoire, autrement dit, elle a utilisé des données qualitatives pour explorer des thèmes qui ont ensuite été vérifiés par les données quantitatives. Le sondage a ainsi été utilisé afin de trianguler les données recueillies lors des entretiens et il a été analysé de façon descriptive. Cette approche consécutive prévue dès le départ a donc facilité l'intégration des données recueillies.

Dwyer a pu constater que les organisations d'entraide ont facilité la réintégration des anciennes personnes détenues et ont promu un sentiment de solidarité basé sur la lutte contre la stigmatisation.

3.5.6 La qualité des études s'est-elle améliorée au fil des ans ?

Une dernière question est pertinente dans cette fin de section : observons-nous une amélioration de la qualité des designs méthodologiques ? La réponse à cette question n'est pas facile et elle peut être biaisée depuis le début, le nombre d'études étant limité par année et par design. Afin de réduire le risque de biais, nous avons pris en compte les études à partir de 2008, éliminé de l'équation les études expérimentales qui sont peu nombreuses et gardé les données des années où il y a eu au moins deux études. La moyenne par design et par année peut être observée dans la Figure 22. Dans tous les cas, l'on peut observer deux étapes : les périodes allant de 2008 à 2014 et de 2015 à 2019. Dans la première étape, les études quantitatives descriptives ont connu une variation importante et ont eu une tendance relative à une diminution de la qualité des méthodes utilisées dans la deuxième période. Ce type d'étude n'est pas conseillé pour une évaluation, notamment de l'impact des programmes, parce qu'elles tracent un portrait de la réalité à un moment donné sans avoir recours à une comparaison temporelle ou de groupes. Les apprentissages que l'on peut tirer de ces études sont ainsi relativement limités. Dans le cas du reste des études, les résultats sont relativement plus élevés dans la première étape. Dans la deuxième période, les résultats ont tendance à s'améliorer (designs quasi expérimentaux) ou à se stabiliser (designs qualitatifs et mixtes). Ces derniers résultats sont plus encourageants parce que c'est la période pendant laquelle les études commencent également à être plus nombreuses.

Figure 22. Évolution de la qualité méthodologique des études par année





3.6 DEUX ÉTUDES DE CAS : L'ÉVALUATION DANS LE CONTEXTE DES PROGRAMMES DE PRÉVENTION DE L'EXTRÊME DROITE ET ÉVALUATION DES PROGRAMMES DE PRÉVENTION EN LIGNE

Avant de présenter les conclusions de ce rapport, nous aimerions aborder deux cas fréquemment évoqués dans la littérature comme étant des défis importants pour l'évaluation des programmes dans le contexte de la prévention de l'extrémisme violent et de la radicalisation violente et qui, comme nous le verrons, nécessitent encore plus d'attention de la part des personnes chercheuses, évaluatrices et preneuses de décision. Nous faisons référence à l'évaluation des programmes de prévention de l'extrémisme violent de droite et des programmes de prévention en ligne.

3.6.1. L'ÉVALUATION DE PROGRAMMES DE PRÉVENTION DE L'EXTRÊME DROITE

L'évaluation des programmes de prévention de l'extrême droite demeure l'un des enjeux majeurs dans ce domaine, notamment en Amérique du Nord et en Europe, où il existe une réelle préoccupation concernant la montée de ces groupes. Comme nous l'avons mentionné dans l'introduction de ce rapport, ces programmes ont été peu évalués, bien que l'une des expériences de réintégration les plus connues (le programme « *EXIT* ») ait démarré bien avant cette nouvelle vague d'extrémismes. En effet, les revues systématiques les plus récentes n'ont trouvé qu'une étude de prévention secondaire (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021) et cinq de prévention tertiaire (Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021) qui visaient ce type d'extrémisme. En revanche, nous avons trouvé 20⁵⁰ études mentionnées dans 17 publications, dont 80 % de ces études en Europe. Il convient de noter que ce chiffre pourrait être encore plus élevé, car certaines de ces études peuvent avoir été exclues de notre revue systématique en raison de la langue dans laquelle elles ont été rédigées, c'est-à-dire en dehors du cadre des études rédigées en anglais, en français ou en espagnol. Cela est particulièrement vrai pour les études qui ont évalué les programmes de prévention de l'extrême droite dans les pays scandinaves, en Allemagne et aux Pays-Bas, où ils ont été mis en œuvre depuis les années 1990.

a. Que savons-nous des programmes évalués ?

À la différence des programmes qui ciblent d'autres types d'extrémisme, ces programmes évalués sont relativement plus focalisés : 60 % des programmes évalués ont ciblé de façon exclusive l'extrême droite, 75 % étaient un projet spécifique, 65 % travaillaient sur la prévention secondaire et 40 % sur la **prévention tertiaire**. Autrement dit, la majorité de ces programmes cible une population à risque. Cette donnée est confirmée par le fait qu'une grande partie de ces programmes cible des individus spécifiques (65 %) et seuls 25 % ciblent un groupe sociétal spécifique (notamment les jeunes) ou la société dans son ensemble.

La plupart des programmes évalués (55 %) se sont concentrés sur le désengagement des individus des groupes ou milieux extrémistes et leur réintégration ultérieure dans la société.

Ces programmes ont été principalement évalués par le biais d'entretiens menés après la mise en œuvre de l'intervention. D'autres méthodes utilisées de manière moins importante dans ce type d'étude ont été celles de l'analyse quantitative descriptive et qualitative des interactions sur Internet (deux études), que nous décrivons un peu plus en détail dans la section suivante, et d'un modèle quasi expérimental. Dans ce dernier cas, l'étude a évalué l'impact d'un programme consistant en une série d'ateliers organisés sur une période de six à huit semaines pour huit groupes de jeunes s'étant désengagés de groupes extrémistes, explorant des domaines tels que l'identité et la culture, la diversité, la migration et l'asile et la compréhension de l'extrémisme (i-works research ltd., 2013). Des sondages ont été administrés avant et après la mise en œuvre de l'intervention.

Un quart des programmes est concentré sur des actions dans des écoles ou en relation avec elles. C'est le cas de l'évaluation d'un programme dans lequel d'anciens extrémistes ont donné des conférences dans les écoles afin de prévenir l'extrémisme violent (Walsh et Gansewig, 2019). Cette initiative a été évaluée par une méthode expérimentale où le groupe d'intervention a été comparé à un groupe témoin. Dans d'autres cas, les méthodes d'évaluation se sont limitées à des entretiens ou des questionnaires appliqués après la mise en œuvre du programme et sur lesquels des analyses principalement descriptives ont été réalisées.

b. Que cherchaient ces évaluations et qui les a réalisées ?

L'objectif de l'évaluation de ces programmes diffère à certains égards des objectifs recherchés par les études qui ont évalué des programmes ciblant d'autres types d'extrémisme (voir section 3.3.1 de ces résultats). Bien que l'impact du programme soit l'objectif d'évaluation le plus mentionné dans toutes les études, cet objectif est proportionnellement plus présent dans les évaluations de ce type de programme, dont 80 % des études mentionnent les avoir réalisées alors que, dans le reste des études, ce pourcentage se situe autour de 70 % en moyenne. Dans le même ordre d'idées, seulement 30 % de ces études évaluatives mentionnent avoir effectué une évaluation de processus alors que, dans le cas des études sur d'autres types d'extrémisme, ces pourcentages atteignent en moyenne 47 % des études. Les évaluations de ce type de programmes semblent ainsi moins ciblées par le processus de mise en place des programmes que les effets qu'ils produisent.

⁵⁰ Nous n'avons considéré ici que les études qui indiquaient clairement qu'elles évaluaient un programme ciblant l'extrême droite. Nous n'avons pas pris en compte les études qui évaluaient des programmes « non spécifiques », bien que l'on puisse considérer que ceux-ci incluent tout type d'extrémisme.

Pour ce qui est du type d'évaluateur ou d'évaluatrice, les analyses démontrent que la majorité des évaluations ont été externes (55 %). Ce pourcentage est toutefois moins important que dans les cas des évaluations de programmes de prévention de l'extrémisme violent dit « djihadiste » ou « islamiste » (71,1 %) et de programmes de prévention non spécifiques (80 %).

c. Comment ces programmes ont-ils été évalués ?

L'approche méthodologique globale a été assez équilibrée, avec 30 % de méthodes qualitatives, 35 % de méthodes quantitatives et 35 % d'études mixtes.

Parmi les études qui ont eu recours à des analyses quantitatives, 7 ont eu une section quantitative descriptive, 5, une section quasi expérimentale et une étude a été exclusivement expérimentale. Bien que le pourcentage d'études qualitatives soit similaire dans toutes les études d'évaluation, le nombre d'études ayant utilisé des techniques quantitatives est supérieur aux évaluations d'autres types de programmes de prévention de l'extrémisme violent. Il s'agit là d'un résultat qui est cohérent avec les efforts mis en place pour évaluer l'impact de ces programmes. Cependant, ces efforts demeurent limités : à peine deux études ont utilisé un groupe témoin et 55 % des évaluations n'a eu lieu qu'une fois le programme terminé. De plus, la plupart des études (70 %) n'utilisaient aucun indicateur mesurant directement la radicalisation, l'extrémisme violent ou les sympathies pour ces phénomènes alors que 80 % utilisaient des indicateurs indirects, c'est-à-dire qui ne mesurent pas directement la radicalisation, l'extrémisme violent ou les sympathies pour ces phénomènes.

d. Quels sont les enjeux de la qualité de ces études ?

Comme dans le cas de la section précédente, la qualité des études a été évaluée par type de design.

La majorité des 14 études qui ont eu recours à des méthodes qualitatives montrent une bonne adéquation du choix de la méthode (92,9 %) et des outils de collecte de données (78,6 %). Cependant, seules deux études ont montré une cohérence suffisante tout au long de la procédure méthodologique et leurs conclusions ont suffisamment été cohérentes par rapport aux données recueillies.

Les sept études qui ont réalisé des analyses quantitatives descriptives montrent des faiblesses semblables à celles déjà analysées dans la section précédente, notamment en ce qui concerne le risque de non-réponse (seuls 14,2 % des études ont eu un taux de réponse acceptable), la stratégie d'échantillonnage (mentionnée dans 42,9 % des études) et la pertinence du type d'analyse statistique utilisée pour répondre aux questions de recherche (42,9 %).

Les cinq études qui ont eu recours à des méthodes quasi expérimentales sont, comparativement, moins robustes que les autres designs et que le même type de design lorsque utilisé dans l'évaluation d'un autre type de programme. Dans aucune étude, par exemple, nous n'avons pu constater que les mesures étaient appropriées ou que les facteurs de confusion ont été pris en considération et, dans seulement une étude, nous avons pu constater la présence du reste des indicateurs.

L'évaluation des études mixtes, pour sa part, concerne davantage l'intégration de deux méthodologies que la performance d'une méthode en particulier. Cette intégration est relative, dépendamment du critère utilisé. L'interprétation des résultats de cette intégration ainsi que le traitement adéquat de leurs divergences sont présents dans cinq des sept études qui ont utilisé ce type de méthodologie. En revanche, nous n'avons trouvé aucune étude qui a bien justifié l'utilisation d'une double méthode, seulement deux études ont bien intégré ces résultats par rapport aux questions de recherche et une seule étude a respecté ces critères de qualité pour chaque méthode.

Si l'augmentation du nombre d'études d'évaluation des programmes de prévention de l'extrémisme violent de droite par rapport aux recensions précédentes peut être un indicateur prometteur, il reste encore un long chemin à parcourir pour déterminer l'utilité de ces programmes. Si ce constat est vrai pour les programmes qui se concentrent sur le désengagement des individus des groupes extrémistes, qui constituent le type le plus courant de programme de prévention de l'extrémisme violent de droite, il est encore plus pertinent pour les programmes qui se concentrent sur d'autres types d'approches, comme les interventions en ligne ou en milieu éducationnel, sur lesquelles nous savons peu de choses. À cela s'ajoutent deux éléments méthodologiques : les difficultés additionnelles de l'évaluation de ce type de programme déjà évoquées tout au long de ce rapport (difficulté d'accès à l'information, enjeux éthiques, etc.) et la faible qualité des évaluations réalisées, que nous avons abordée dans la section précédente. De futures études intégrant des évaluations réalisées dans les pays scandinaves, aux Pays-Bas et en Allemagne pourraient apporter un éclairage supplémentaire sur les questions qui demeurent ouvertes ici.

3.6.2 ÉVALUATION DES PROGRAMMES DE PRÉVENTION EN LIGNE

Les programmes de prévention en ligne prennent de plus en plus d'importance dans le domaine de la prévention et dans d'autres domaines connexes, c'est pourquoi nous avons décidé de leur consacrer une section à part où nous pouvons présenter les résultats spécifiques des designs utilisés et les défis de l'évaluation dans ce domaine.

Comme pour les autres types de programmes de prévention de l'extrémisme violent, les évaluations de ces programmes sont rares. Nous avons identifié seulement 14 études d'évaluation de programmes en ligne (6,4 %), 2 études qui ont évalué un programme mixte combinant une composante en ligne et hors ligne (0,9 %), 7 études qui ont évalué un programme qui mentionnait avoir une composante en ligne, mais qui n'a pas été évalué (3,2 %) et 4 études qui ont évalué des programmes qui étaient hors ligne, mais qui ciblaient l'espace numérique (alphabétisation numérique = 1,8 %). Nous nous concentrons principalement sur les études qui ont évalué des programmes en ligne et sur l'évaluation de programmes mixtes (16 études). Toutefois, dans certains cas, nous faisons également référence à des programmes hors ligne qui ont abordé l'espace numérique.

a. Que savons-nous des programmes évalués ?

La majorité de ces études ont évalué un programme en ligne de contre-narratifs (13 études) qui consistaient principalement en des campagnes de sensibilisation pour la population cible. C'est le cas, par exemple, du programme « Extreme dialogue » au Canada qui a duré 16 mois et produit une série de films de sensibilisation diffusés sur Internet et visant à réduire l'attrait de l'extrémisme chez les jeunes et à offrir une alternative positive à la quantité croissante de contenus extrémistes sur Internet (SecDev.Foundation, 2016). La procédure d'évaluation de ces campagnes de sensibilisation suit presque exclusivement le même schéma : mesurer la portée de la campagne (combien de personnes ont regardé les vidéos ou ont été exposées au contenu) et son impact en analysant les différentes interactions générées par le contenu diffusé (nombre de fois où le contenu a été partagé, nombre de « j'aime » et l'analyse des commentaires, etc.). Ce sont donc des évaluations qui ont utilisé les données recueillies sur Internet. Deux autres études ont évalué la même campagne en recueillant les données directement sur deux échantillons différents : un groupe d'étudiants collégiaux américains et un groupe d'étudiants somalis américains (McDowell-Smith et al., 2017 ; Speckhard et al., 2019). Dans ce cas, les auteurs ont testé deux vidéos (« A Sex Slave for You—A gift from Abu Bakr al-Baghdadi » et « Rewards for Joining the Islamic State ») qui allaient se diffuser par la suite en ligne à travers l'analyse quantitative des propos des étudiantes et étudiants qui ont participé à des groupes de discussion après l'exposition à l'intervention. Une seule étude a également mené des entretiens et des groupes de discussion avec des jeunes et des membres du programme pour connaître plus en détail les impacts produits par ces programmes (Jailobaeva et Asilbekova, 2017).

Deux autres études ont évalué deux programmes d'intervention directe sur les médias sociaux auprès de personnes considérées comme à risque ou ayant un profil extrémiste (Davey et al., 2018 ; Frenett et Dow, 2015). Ces deux programmes ont été conçus, mis en œuvre et évalués par l'« Institute for Strategic Dialogue » (ISD) et consistaient principalement en une conversation directe, personnalisée et privée avec un interlocuteur (généralement un ancien extrémiste) dans le but de déradicaliser ou de désengager la personne ciblée d'un mouvement extrémiste ou de la dissuader de partager ou de consommer des contenus extrémistes. La procédure d'évaluation de ces programmes était également la même : le taux de réponse et la durabilité des échanges (conversations avec cinq messages ou plus entre la personne candidate et la personne interlocutrice) ainsi que l'analyse du contenu des messages et des indices d'un éventuel impact positif au cours des conversations. Dans les deux cas, les interactions étaient limitées et le nombre restreint d'interactions a empêché une évaluation efficace de l'impact de ces programmes. Ces évaluations ont également eu tendance à considérer tout type d'interaction comme positif. Dans le cas de l'étude de Frenett et Dow (2015), par exemple, cela incluait des changements de comportement tels que la modification des paramètres de confidentialité du compte utilisateur ou le blocage de la personne interlocutrice. Le nombre d'interactions pour mesurer la durabilité était réduit à cinq qui pouvaient être les cinq premiers

échanges pour commencer la conversation. Finalement, une dernière étude a évalué une politique publique nationale par le biais d'une analyse documentaire et d'entretiens avec des fonctionnaires (Warrington, 2018).

Les programmes hors ligne qui visaient l'espace numérique (4) ont principalement visé à développer l'esprit critique et la littératie numérique (Colibaba et al., 2017 ; Gatewood et Boyer, 2019 ; Parker et al., 2018 ; Reynolds et Parker, 2018). Le programme « Digital Resilience » a, par exemple, visé à fournir aux jeunes les connaissances, les compétences, les attitudes et les comportements dont ils ont besoin pour devenir de bons citoyens numériques au XXI^e siècle, en se concentrant sur les défis de l'extrémisme violent en ligne ainsi que sur la gestion efficace des discours de haine et l'identification de la désinformation active (Reynolds et Parker, 2018). Les évaluations de ce type de programme ont tendance à être plus complexes que celles mentionnées ci-dessus, impliquant en grande partie des groupes de discussion et l'utilisation de sondages et des mesures répétées avant et après l'intervention.

En ne considérant cette fois que les programmes en ligne ou les programmes ayant une composante en ligne évaluée, ceux-ci sont principalement situés en Amérique du Nord (6) et en Asie (4), le reste étant situé en Afrique (2) et en Europe (1). Les trois autres ne visaient pas un pays ou une région spécifique. La répartition par pays est très diverse, les États-Unis (4) et le Canada (2), les deux ayant le nombre le plus important de ces programmes évalués. Dans le cas de l'Asie, les programmes ont été mis en œuvre en Indonésie, en Irak, au Kirghizstan et au Pakistan. Ces programmes en ligne évalués ont été conçus pour fonctionner à différents niveaux de prévention, étant ancrés aux niveaux secondaire (6) et tertiaire (5) ainsi qu'au niveau primaire (5). La majorité de ces programmes ciblent l'extrémisme islamiste (11), suivis par les programmes ciblant l'extrême droite (4) et les programmes non spécifiques (ciblant plus d'un type d'extrémisme) (4).

b. Que cherchaient ces évaluations et qui les a réalisées ?

Les études qui ont évalué ces programmes sont relativement récentes, la première ayant été publiée en 2015, et la plupart des autres, entre 2016 et 2017. L'objectif de la majorité de ces études était d'évaluer l'impact du programme en question (81,3 %) ; deux études ont tenté d'évaluer les processus d'implémentation. La majorité des évaluations ont été internes (11), c'est-à-dire réalisées par les équipes qui ont conçu ou mis en œuvre le programme.

c. Comment ces programmes ont-ils été évalués ?

En termes strictement méthodologiques, ce type d'étude a principalement analysé les informations disponibles sur les réseaux sociaux (métrique) d'un point de vue quantitatif descriptif (93,8 %), en incorporant quelques éléments d'analyse qualitative, notamment en ce qui concerne les commentaires laissés sous les publications de la campagne ou à partir des interactions dans le cas des interventions directes. Ce type d'étude, comme indiqué ci-dessus, utilise rarement les outils traditionnels de collecte de données tels que les sondages (4 études), les groupes de discussion (4 études) ou les entretiens (5 études). Les données qualitatives recueillies, par exemple, sont souvent chiffrées afin d'avoir un aperçu quantitatif de l'évaluation. Toutes les études ont utilisé des indicateurs indirects et seules quatre études ont utilisé des indicateurs mesurant directement l'extrémisme violent ou une variable connexe. Les designs ne sont pas non plus très sophistiqués, réalisant la collecte et l'analyse après la fin des interventions, sans groupes témoins et se concentrant sur la description des informations collectées. La transparence n'est pas non plus un point fort, car seules cinq études mentionnent clairement le nombre de personnes participant à ces évaluations. Malgré le fait que la quasi-totalité de ces études cherchait à mesurer les effets des programmes, et malgré l'utilisation d'analyses quantitatives, ces études n'ont pas utilisé de mesures d'association entre l'intervention et ses effets.

d. Quels sont les enjeux de la qualité de ces études ?

Le MMAT est plus pertinent pour les études traditionnelles dans lesquelles le nombre de participantes et participants et les outils de collecte de données peuvent être identifiés. Ce modèle traditionnel a rarement été suivi pour les évaluations de programmes en ligne. Cependant, certains éléments peuvent être extraits des données collectées. Par exemple, 14 des 16 études ont analysé les données d'un point de vue qualitatif, 15 études ont eu recours à des méthodes quantitatives et 13 études ont été mixtes. Dans ce cas, l'approche qualitative et les méthodes de collecte de données ont été, pour la plupart, adéquates pour répondre aux questions de recherche. En revanche, bien qu'environ la moitié des études ait appuyé les interprétations sur des données, la cohérence de l'ensemble de la procédure était faible, tout comme la cohérence entre ce qui était énoncé dans les conclusions et les données recueillies. Aucune des 15 études qui ont eu recours à des méthodes quantitatives n'a utilisé de designs expérimentaux ou quasi expérimentaux, utilisant plutôt, comme indiqué ci-dessus, des analyses quantitatives-descriptives. Le principal risque de ce type d'étude tient au fait que la participation est faible et que la base de l'analyse n'est pas constituée d'individus, mais d'actions sur les médias sociaux, comme les vues, le partage de contenu, etc. Ainsi, dans certains cas, aucune étude ne remplissait les critères nécessaires, par exemple, le risque de biais de non-réponse, et très peu rencontraient le reste des indicateurs (pertinence de l'analyse statistique et de la stratégie d'échantillonnage). La représentativité des individus sélectionnés a été le seul indicateur pour lequel la plupart des études ont satisfait aux critères, bien qu'il faille noter que ces critères, comme indiqué ci-dessus, sont assez flexibles. L'intégration des sections qualitatives et quantitatives dans les études mixtes a fait l'objet d'une évaluation négative. Sur les 13 études qui ont utilisé les deux méthodologies, aucune n'a justifié l'utilisation des deux modèles et très peu ont été capables d'intégrer efficacement les deux designs. Seuls 66 % des études ont pu interpréter correctement les résultats de cette intégration et 54 % seulement ont pu intégrer les divergences entre les deux.

En conclusion, malgré l'importance d'Internet dans le processus de radicalisation et dans l'extrémisme violent et les efforts considérables déployés pour mettre en œuvre des programmes de prévention en ligne, les évaluations existantes dans ce domaine n'en sont probablement qu'à leurs débuts et on ne peut tirer que très peu de conclusions de leurs résultats.

La plupart de ces études ont accordé une grande importance aux informations facilement disponibles sur Internet, telle que les métriques, et peu d'importance à l'évaluation cohérente de l'impact de ce type de programme sur les attitudes, les émotions et les comportements des individus qu'il cible. Les interprétations de l'analyse des indicateurs associés à la « métrique » des réseaux sociaux sont aussi très facilement biaisées. La réaction à un post ou à une vidéo par intermédiaire d'un « j'aime » peut avoir des interprétations multiples. Sur Facebook, par exemple, il n'est pas rare de voir une publication sur une personne décédée avec des « likes » et non des « tristes ». Dans ce contexte, le « like » peut être interprété non pas comme une appréciation positive de la nouvelle, mais plutôt comme un signe de soutien. Alors, comment évaluer un programme sur la base d'indicateurs qui peuvent être interprétés de multiples façons ? Cela nous amène à la question de la validité méthodologique de ces évaluations. Mesurent-elles vraiment ce qu'elles sont censées mesurer ?

Ces évaluations n'ont pas non plus accordé beaucoup d'attention aux facteurs de mise en œuvre qui pourraient faciliter ou entraver le processus d'intervention, du moins pas de manière systématique. Lorsque ces facteurs sont évoqués, ils apparaissent davantage comme des leçons tirées de la mise en œuvre et de l'expérience des personnes conceptrices-évaluatrices que comme le résultat d'une analyse systématique d'informations provenant de différentes sources.

Un dernier élément à prendre en compte concerne l'éthique de l'évaluation, notamment dans le cas d'interventions directes en ligne. Dans les deux programmes qui ont réalisé ce type d'intervention, il a été indiqué que les informations résultant des échanges avec les interlocutrices et interlocuteurs ont été analysées. Toutefois, il n'est pas précisé si ces informations partagées à titre privé ont reçu le consentement des personnes concernées pour être analysées.

Recommandations

Chaque public cible sera illustré par les icônes suivantes :



Chargées et chargés de programmes



Bailleurs de fonds



Actrices et acteurs gouvernementaux



Évaluatrices et évaluateurs



Conceptrices et concepteurs de programmes



Intervenantes et intervenants



Usagères et usagers des programmes

1.1 CONSIDÉRATIONS GÉNÉRALES

- a) Développez et encouragez une culture d'évaluation au sein des organisations œuvrant dans le domaine de la prévention de la radicalisation et de l'extrémisme violent.
- b) Favorisez une évaluation indépendante et de meilleure qualité dans le cas des programmes de prévention de l'extrême droite et de programmes de prévention en ligne.

1.2 FINANCEMENT

- a) Le financement de tout programme en prévention devrait comprendre, dès le début, une ligne budgétaire spécifique allouée au volet d'évaluation (minimum 10 %).
- b) Cette ligne budgétaire doit être suffisante pour assurer une évaluation de qualité, autrement dit, qui permet d'assurer l'engagement du personnel spécialisé en évaluation et en prévention, l'acquisition du matériel nécessaire pour la cueillette et l'analyse de données, les coûts des déplacements et pour assurer le temps nécessaire pour bien mener cette évaluation. La durée de l'évaluation doit s'estimer en fonction de la taille du programme, du nombre de ses composantes et de sa complexité ainsi qu'en fonction de son étendue géographique.

1.3 ÉQUIPE D'ÉVALUATION

- a) Composez une équipe d'évaluation diversifiée et représentative au regard du programme et des actrices et acteurs impliqués, qui inclut autant des membres externes pour assurer l'indépendance de l'évaluation qu'internes afin d'assurer une connaissance approfondie du programme. Dans ce dernier cas, favorisez l'inclusion de personnes intervenantes dans l'équipe d'évaluation.
- b) Favorisez l'évaluation réalisée par des équipes locales et, dans le cas où cette évaluation est réalisée par une équipe internationale, favorisez l'inclusion d'évaluatrices et évaluateurs locaux dans l'équipe principale.



- c) Impliquez les personnes évaluatrices dès la conception ou la mise en œuvre du programme.
- d) Assurez-vous que la personne évaluatrice parle la langue du pays dans lequel l'évaluation est menée et qu'elle a une connaissance approfondie du contexte local.

1.4 CONCEPTUALISATION DU PROTOCOLE D'ÉVALUATION



- a) Utilisez une théorie du changement claire qui se traduit par des objectifs concrets et spécifiques à partir desquels les personnes évaluatrices peuvent déterminer les indicateurs de changement à évaluer.



- b) Favorisez la cohérence tout au long du processus d'évaluation afin que les résultats obtenus soient mesurés au moyen d'indicateurs appropriés qui correspondent aux objectifs de l'évaluation et à la méthodologie utilisée.



- c) Travaillez avec un horizon temporel adéquat pour collecter et analyser des informations suffisantes. Ce calendrier doit considérer la reconstruction de la théorie du changement, les actions avant et après l'intervention (le cas échéant), l'analyse des données, la rédaction du rapport et les actions de mobilisation des connaissances.



- d) Assurez-vous de documenter les informations relatives au programme depuis le début des actions afin que les personnes évaluatrices puissent compter sur une source d'information riche et officielle.



- e) Dans le cas de l'utilisation d'indicateurs indirects (qui ne mesurent pas directement la radicalisation, l'extrémisme violent ou les sympathies pour ces phénomènes), décrivez toujours la relation entre ces indicateurs et l'extrémisme violent.



- f) Assurez-vous toujours de concevoir, dès le début, un plan de mobilisation de connaissances diversifié qui inclut le rapport, un sommaire exécutif et, du moins, un atelier ou une présentation à l'équipe du programme.

1.5 MÉTHODOLOGIE



- a) Favorisez l'utilisation de devis mixtes (quantitatifs et qualitatifs) pour l'évaluation de programmes complexes puisqu'ils facilitent autant l'évaluation de l'impact que l'évaluation du processus.



- b) Encouragez les évaluations de processus pour avoir davantage de connaissances sur les facteurs qui exercent une influence sur la mise en place de programmes.



- c) Favorisez les devis à mesures répétées dans les évaluations de ces programmes, c'est-à-dire basés sur plusieurs observations des mêmes sujets à au moins deux moments différents.



- d) Évitez le recours à des devis quantitatifs descriptifs comme seule méthode à utiliser.



- e) Ne multipliez pas le nombre d'outils d'évaluation à remplir, car ceci peut diminuer le taux de réponse des personnes actrices clés et peut également nuire à votre capacité d'analyse ou à la profondeur de cette dernière en raison d'un nombre trop élevé de données.



- f) Dans le cas de devis quantitatifs, utilisez des échantillons avec un nombre suffisant de participantes et participants pour permettre les analyses statistiques nécessaires et appropriées pour répondre à la question d'évaluation.



g) Dans le cas de devis expérimentaux, assurez-vous que la randomisation est bien décrite et que les groupes témoins et de traitement sont comparables, dès le départ.



h) Dans le cas de devis quasi expérimentaux, assurez-vous que les échantillons représentent la population visée et tenez compte des facteurs de confusion qui peuvent influencer les résultats.



i) Dans le cas de devis qualitatifs, assurez-vous de disposer d'un échantillon diversifié de personnes actrices et usagères à interroger, tant à l'extérieur qu'à l'intérieur du programme.



j) Dans le cas de devis qualitatifs, veillez à ce que l'interprétation des données soit appuyée par des informations concrètes.



k) Dans le cas de devis mixtes, assurez-vous qu'il existe une bonne intégration entre les composantes quantitatives et qualitatives des études.



l) Dans le cas d'évaluations d'impact, évitez d'utiliser des modèles expérimentaux pour évaluer des programmes de prévention tertiaire.



m) Dans le cas d'évaluations d'impact, favorisez l'utilisation de devis expérimentaux et quasi expérimentaux avec un groupe de comparaison, lorsque les conditions éthiques et du contexte du programme le permettent.

1.6 DURANT L'ÉVALUATION



a) Assurez-vous que l'évaluation se réalise dans un cadre sécurisant pour les personnes participantes de l'étude et les évaluatrices et évaluateurs.



b) Favorisez toujours la création d'un lien de confiance entre la personne évaluatrice et les personnes participantes de l'étude afin d'obtenir des informations de confiance.



c) Utilisez judicieusement le temps consacré à la participation des personnes intervenantes à l'évaluation, car une participation excessive peut nuire à leur engagement.



d) Favorisez un accès direct des personnes évaluatrices aux participantes et participants et aux informations pertinentes pour l'évaluation.

1.7 POST-ÉVALUATION



a) Rédigez les rapports d'évaluation dans la langue d'usage du pays où l'évaluation a été réalisée.



b) Soyez transparent quant à la méthodologie utilisée et décrivez-la de façon détaillée.



c) Décrivez toujours les limites et les conflits d'intérêts des études et, lorsqu'ils n'existent pas, mentionnez toujours cette situation.



d) Décrivez toujours les participantes et participants à l'étude évaluative.



e) Assurez-vous toujours de mettre en place un plan de mobilisation de connaissances diversifié qui inclut le rapport, un sommaire exécutif, un atelier ou une présentation à l'équipe du programme.



Conclusion

L'évaluation dans ce domaine est difficile, certes, mais possible. Une grande partie de la littérature a jusqu'à présent principalement exposé les difficultés méthodologiques et pratiques de l'évaluation des programmes dans le domaine de la prévention de l'extrémisme violent. Des difficultés d'une telle nature que l'évaluation semblait être un problème insoluble, plein de pièges et d'impasses. L'idée du caractère exceptionnel du phénomène de l'extrémisme violent, qui a exigé à son tour des mesures aussi exceptionnelles dans tous les espaces, y compris l'évaluation, y a probablement contribué. L'approche pragmatique que nous avons évoquée dans la partie finale de la section 1 de ce rapport résume en quelque sorte cette perception : nous devons évaluer ce que nous pouvons avec les données dont nous disposons et avec des méthodologies *ad hoc*. Cependant, l'idée du caractère exceptionnel du phénomène commence à se dissiper. Les facteurs qui expliquent l'extrémisme violent ne sont pas très différents de ceux que l'on trouve dans des domaines connexes, telle que l'étude de la délinquance (Wolfowicz et al., 2019), et les programmes de prévention ne sont pas non plus très différents des programmes de prévention dans d'autres domaines de la violence (Madriaza et Ponsot, 2015). **L'évaluation de ces programmes, y compris les méthodes et les objectifs, comme nous l'avons vue tout au long de ce rapport, n'est pas non plus, à quelques exceptions près, très différente de l'évaluation d'autres programmes complexes de prévention de la violence.**

Le nombre limité d'évaluations, pendant les années précédentes et dans les recensions passées, s'explique davantage par le caractère récent du concept de radicalisation menant à la violence et d'extrémisme violent et de sa prévention en tant que champs d'études et par l'urgence d'agir de ces quinze dernières années, qui ont multiplié les actions publiques pour lutter contre un phénomène qui semblait déborder. Un biais de publication peut également y avoir contribué, car une grande partie des analyses de la littérature passée s'est concentrée principalement sur la littérature académique. La littérature grise, en revanche, s'est révélée être une source d'information extrêmement riche. Ce que nous observons alors, c'est plutôt une mise à jour de l'évaluation et un tournant dans le domaine, notamment depuis 2016, où l'urgence d'intervenir à tout prix fait place à une intervention susceptible d'être évaluée et

donc, à un plus grand intérêt pour la production de preuves et de connaissances émergeant de la pratique. Les 219 études identifiées en sont la preuve. Notre stratégie de recherche peut également y avoir contribué. Nous avons recherché toutes les études qui ont fait une évaluation d'un programme particulier, indépendamment de la qualité des méthodes utilisées. L'objectif était de brosser un portrait de la réalité de l'évaluation de nos jours dans ce domaine. Une revue systématique qui se concentrerait sur les effets de ces programmes exclurait probablement un grand pourcentage des études ici retenues.

Malgré ces constats, l'évaluation dans ce domaine continue de présenter certaines particularités qu'il convient d'aborder. Mesurer directement l'extrémisme violent demeure, par exemple, extrêmement difficile en raison des problèmes de définition et d'opérationnalisation du phénomène. Les évaluatrices et évaluateurs ont plutôt utilisé des indicateurs indirects qui sont, sans aucun doute, plus appropriés au contexte du programme en question, un fait qui nous invite paradoxalement à réfléchir à la particularité de ce type de programme qui cherche à influencer un phénomène difficilement mesurable. La relation directe entre ces indicateurs et l'extrémisme violent dans le cas de ces programmes ne peut donc être que théorique.

Un autre élément intéressant que nous avons pu constater dans ce rapport est l'émergence d'études expérimentales pour l'évaluation des effets de l'intervention qui n'avaient pas été identifiées jusqu'à présent. Cependant, leur mise en œuvre semble être plus facile dans le cas de programmes de prévention plus universels tels que les programmes de prévention primaire et primaire ciblée. Dans le cas de programmes plus spécifiques, tels que les programmes de prévention secondaire et tertiaire, les enjeux éthiques ainsi que d'autres questions plus pratiques réduisent les possibilités d'utilisation de l'expérimentation comme standard dans ce domaine. En effet, à mesure que les programmes deviennent plus spécifiques et se concentrent sur une population plus restreinte, l'accès aux sources d'information et aux personnes participantes est fortement réduit. Les études quasi expérimentales, notamment celles qui comportent des mesures pré et post-intervention, apparaissent donc comme une alternative viable et plus sophistiquée pour

mesurer les effets d'une intervention dans ce domaine et sont présentes à tous les niveaux de la prévention. Une autre méthodologie qui commence à devenir une norme dans le domaine est celle des études mixtes. Ces modèles nous permettent de quantifier et de mesurer de manière cohérente les effets de l'intervention et fournissent en même temps des informations pour l'évaluation de la mise en œuvre du programme. Comme nous l'avons vu dans l'étude internationale (Madriaza et al., 2021), où nous avons interrogé des personnes évaluatrices et intervenantes, les études qualitatives ne peuvent être négligées, car elles fournissent des informations et des contextualisations qui ont beaucoup plus de sens pour la pratique elle-même et, en particulier, pour les intervenantes et intervenants. Les études mixtes incluant des modèles quasi expérimentaux peuvent constituer une solution concrète pour améliorer la qualité des évaluations sur le terrain.

Bien que la présence de designs quantitatifs plus complexes soit prometteuse dans le cas de l'évaluation de l'impact des programmes de prévention, un fait manifestement important est que, en particulier dans ce domaine, il existe une certaine confusion entre ce qu'est l'évaluation d'impact et l'évaluation en général. En d'autres termes, on a tendance à aborder les évaluations comme si toute évaluation était une évaluation des effets des programmes. Comme nous l'avons vu dans notre rapport, pour une majorité des études, l'évaluation des effets du programme apparaît comme l'objectif le plus fréquent utilisé pour justifier l'évaluation. Cependant, dans une grande partie de ces études, en particulier celles qui ont utilisé des modèles quantitatifs descriptifs, l'intention d'évaluer les effets ne correspond pas à l'utilisation de méthodologies appropriées pour les évaluer. Nous nous retrouvons alors avec des études qui se sont plutôt concentrées sur la description quantitative des activités, sur la description de la satisfaction des bénéficiaires ou sur la description d'autres variables associées qui peuvent difficilement rendre compte de l'impact que le programme a eu sur ses usagères et usagers. Ce biais peut également être observé dans le cas des recensions passées qui ont limité la compréhension de l'évaluation des programmes à la question de savoir si les programmes ont eu les résultats escomptés. Alors que ces informations sont essentielles dans le domaine et pour les politiques publiques, peu de recensions ont exploré les facteurs qui expliquent pourquoi un programme peut ou non réussir. Les évaluations de processus sont la clé de la compréhension de ces mécanismes, mais elles n'ont pas fait l'objet de recensions spécifiques à ce jour pour comprendre quels facteurs de mise en œuvre peuvent favoriser ou entraver la réussite d'un programme. Les études de Gielen (2017) et Veldhuis (2015) ont apporté un éclairage dans ce sens, mais sont encore insuffisantes pour comprendre les mécanismes qui influencent ce processus.

Dans une autre optique, l'analyse de la qualité des méthodes utilisées nous a également donné quelques indices sur les défis que l'évaluation doit relever dans

ce domaine. La première conclusion est que très peu d'études ont satisfait à tous les critères mesurés par le MMAT et, qu'en moyenne, les résultats de chacune des méthodes évaluées sont modérés, les études quantitatives descriptives étant celles qui ont eu une évaluation plus faible. Ce dernier point confirme l'idée que ce type d'études doit être évité dans l'évaluation de programmes, car le design est moins adapté pour évaluer les changements induits et parce que la qualité de ceux-ci est moins significative. Cependant, chacune des méthodes analysées présente des points forts et des points faibles. Dans le cas des études qualitatives, bien qu'il s'agisse de l'une des méthodes les mieux évaluées, il existe des faiblesses dans la relation entre les données et l'interprétation des résultats qui n'est pas suffisamment corroborée par des informations concrètes. Cela peut conduire à exacerber les biais subjectifs qui sont déjà présents dans le design qualitatif en général. Le principal problème des études descriptives quantitatives est la transparence des méthodes utilisées. La plupart de ces études ne disposaient pas d'informations suffisantes pour que nous puissions effectuer l'évaluation de la qualité de l'étude dans son ensemble ou les informations étaient très incomplètes. Les designs expérimentaux montrent une meilleure performance en général, mais présentent des problèmes dans la description de l'assignation aléatoire entre le groupe témoin et le groupe de traitement et dans la comparabilité des groupes au début, deux aspects qui sont fondamentaux pour la bonne performance des expériences. Le degré d'adhésion des participantes et participants et le degré d'exhaustivité des données collectées montrent cependant une bonne performance dans ces études. Les études quasi expérimentales (celles qui présentent presque tous les éléments des études expérimentales à l'exception de la distribution aléatoire des personnes participantes) connaissent un développement positif ces dernières années. L'un des points à améliorer concerne la représentativité des échantillons choisis et l'importance accordée aux facteurs de confusion. Ces facteurs peuvent altérer l'association entre l'intervention et ses effets et sont donc essentiels pour savoir si l'intervention a effectivement eu un impact sur les personnes qui ont participé. Enfin, dans le cas des études mixtes, un problème fréquent qui apparaît est l'intégration entre les composantes quantitatives et qualitatives de l'étude. Malgré ce problème, nos résultats nous suggèrent que ce type d'évaluation peut devenir la norme dans le domaine. Le problème le plus transversal des études évaluatives identifiées en termes de qualité est peut-être les limites dans la transparence des méthodes, c'est-à-dire la quantité d'informations et le niveau de détail de ces informations dans les sections méthodologiques. Or, la transparence est essentielle pour pouvoir déterminer si l'on peut se fier ou non aux résultats présentés.

Le niveau de prévention, et donc le degré de spécificité ou d'universalité du programme évalué, comme nous l'avons vu précédemment, semble exercer une influence importante sur la manière dont ces programmes sont évalués. À différents moments du rapport, nous avons

pu observer cette différence. L'évaluation d'impact, par exemple, est proportionnellement plus présente parmi les évaluations de programmes plus universels que l'évaluation de processus. Dans le cas des programmes de prévention tertiaire, cette proportion favorise les évaluations de processus. Les évaluations de programmes plus universels ont également une plus grande présence d'évaluateurs et d'évaluatrices externes, une plus grande proportion d'études qui prennent des mesures pré et post-intervention et une meilleure évaluation de la qualité méthodologique dans le cas des études qualitatives et quasi expérimentales. Dans les programmes de prévention plus ciblés, en revanche, des indicateurs plus directs sont utilisés, mais les limites de l'évaluation sont moins bien indiquées. L'évaluation de ce type de programme fait donc face à des défis différents, selon le niveau d'universalité ou de spécificité du programme, et peut donc nécessiter des modèles d'évaluation spécifiques à la réalité de chaque contexte.

Le type d'extrémisme vers lequel le programme de prévention évalué est orienté présente également des résultats à prendre en compte. Tout d'abord, les programmes de prévention de l'extrémisme violent dit djihadiste ou islamiste, continuent de faire l'objet d'un nombre important d'évaluations, mais ce sont surtout les programmes non spécifiques qui affichent une croissance significative au cours des dernières années couvertes par cette étude. Il est très probable que cette tendance soit également observée dans le cas des programmes qui n'ont pas été évalués et qui s'explique principalement par les évaluations négatives des programmes qui ont exclusivement ciblé l'islamisme radical. En revanche, il y a encore relativement peu de programmes de prévention de l'extrémisme violent de droite qui ont été évalués, même si nous constatons une légère tendance à la hausse, et il y a encore relativement peu de programmes de prévention de l'extrémisme violent de droite qui ont été évalués, et il y a encore moins de programmes évalués ciblant l'extrême gauche. Ces analyses, comme nous l'avons évoqué à plusieurs reprises, peuvent être biaisées puisque nous n'avons pas pu accéder à des études dans d'autres langues où ce type d'étude peut être plus répandu, notamment en ce qui a trait à la prévention de l'extrême droite en Europe du Nord. L'évaluation des programmes de prévention de l'extrémisme violent de droite reste donc l'un des principaux enjeux dans le domaine de la prévention de l'extrémisme violent, surtout si l'on considère l'inquiétude croissante que suscitent ces groupes, notamment en Amérique du Nord et en Europe, ainsi que les discours haineux qu'ils génèrent. D'un point de vue méthodologique, les études évaluatives de ces programmes évaluent paradoxalement les effets de ces programmes sur leurs utilisatrices et utilisateurs par des méthodes qualitatives, alors que les méthodes quantitatives sont probablement plus pertinentes pour cet objectif d'évaluation. En outre, très peu d'entre elles comportent des groupes de comparaison, la majorité n'évalue le programme qu'après sa fin et la qualité des

méthodes utilisées est systématiquement jugée inférieure à celle des autres types d'études. Il y a certainement encore beaucoup de chemin à parcourir pour évaluer davantage de programmes de ce type et pour disposer d'évaluations de meilleure qualité.

Bien que le nombre d'évaluations soit beaucoup plus élevé en Europe, l'une des conclusions de cette étude est la présence d'un nombre important d'études évaluatives dans d'autres régions du monde, notamment en Afrique et en Asie, et le faible nombre en Amérique du Nord, malgré le fait que cette dernière région ait une présence importante d'évaluations dans des domaines connexes, comme la prévention du crime. Le cas du Canada est particulièrement préoccupant, car il compte très peu d'études évaluatives publiées. Ces chiffres absolus cachent toutefois un certain nombre de particularités. Premièrement, les évaluations sont concentrées dans quelques pays seulement : presque toutes les études nord-américaines proviennent des États-Unis, la moitié des études européennes proviennent du Royaume-Uni, un tiers des études asiatiques proviennent d'Indonésie et un cinquième des études africaines a été réalisé au Kenya. Deuxièmement, malgré cet équilibre relatif entre certaines régions, la prédominance de l'approche occidentale est écrasante dans la plupart d'entre elles. Le cas de l'Afrique est peut-être le plus évident. On ne trouve guère d'auteurs et d'auteurs africains parmi les études qui ont évalué ce type de programme sur le continent, la majorité des chercheuses et chercheurs étant originaires des États-Unis et, dans de nombreux cas, ne parlent pas la langue du pays dans lequel ils et elles ont mené l'évaluation, comme cela a été noté dans l'analyse des limites des études retenues. En fait, nous n'avons trouvé que 15 études aux États-Unis et 95 auteurs et auteures au total provenant de ce même pays. La langue des rapports d'évaluation pour le cas africain était exclusivement l'anglais, malgré le fait que le français était parlé dans de nombreux cas, dans les pays où le programme a été mis en œuvre. Dans plusieurs sections de ce rapport, nous nous sommes posés comme question « Pour qui ces évaluations sont-elles réalisées ? ». Le but de l'évaluation n'est pas de rédiger un rapport, mais de fournir un ensemble de connaissances qui permettront de prendre de meilleures décisions et, certainement, d'améliorer les programmes mis en œuvre. Nous ne savons pas si d'autres formes de mobilisation des connaissances ont été mises en œuvre pour fournir un retour d'information aux équipes de terrain, mais au moins, dans le cas particulier des études africaines, l'évaluation semble être davantage une nécessité administrative pour les organismes de financement qu'un réel intérêt à produire de meilleures interventions fondées sur des preuves.



Références

Abu-Nimer, M., et Nasser, I. (2017). Building peace education in the Islamic educational context. *International Review of Education*, 63(2), 153-167. <https://doi.org/10.1007/s11159-017-9632-7>

Admo, N., Wood, A., et Ducol, B. (2018). *Une place de choix pour dire et se dire, Évaluation d'implantation et d'impacts d'un projet de prévention de la radicalisation menant à la violence* (p. 54). Collège de Maisonneuve, Centre de prévention de la radicalisation menant à la violence et Institut Pacifique. https://info-radical.org/wp-content/uploads/2018/10/UNE_PLACE_DE_CHOIX_POUR_DIRE_ET_SE_DIRE.pdf

Albert, A., Cabalion, J., et Cohen, V. (2020). *Un impossible travail de déradicalisation*. Erès Editions.

Aldrich, D. P. (2012). Radio as the Voice of God : Peace and Tolerance Radio Programming's Impact on Norms. *Perspectives on Terrorism*, 6(6), 34-60.

Aldrich, D. P. (2014). First steps towards hearts and minds ? USAID's countering violent extremism policies in Africa. *Terrorism and Political Violence*, 26(3), 523-546. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1080/09546553.2012.738263>

Algristian, H., Choiriya, D. D., Abdillah, D. S., Ulya, A., Sodali, H. A., Muhammad, A. R., et Handayani, H. (2019). Why does de-radicalization seem a utopia ? Evaluation on "Children of the Country" program. *Journal of Public Health in Africa*, 10, 148-151. <https://doi.org/10.4081/jphia.2019.1211>

Ali, Y., et Saragih, H. J. R. (2018). *Implementation of Contra-Radicalization in Alkhairaat Educational Institutions*. 810-817.

Al-Maqosi, Y. A., Al-Bataineh, M. T., et Al-Kilani, A. M. (2019). The Effectiveness of an Educational Program for Developing Tolerance Values and Resistance to Intellectual Extremism at Secondary Level in Jordan. *Journal of Educational and Psychological Studies [JEPS]*, 13(4), 628-642. <https://doi.org/10.24200/jeps.vol13iss4pp628-642>

Aly, A., Taylor, E., et Karnovsky, S. (2014). Moral disengagement and building resilience to violent extremism : An education intervention. *Studies in Conflict and Terrorism*, 37(4), 369-385. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1080/1057610X.2014.879379>

Amanullah, Z., et Harrasy, A. (2017). *Between Two Extremes, Responding to Islamist and tribalist messaging online in Kenya during the 2017 elections* (p. 24). Institute for Strategic Dialogue (ISD). <https://www.isdglobal.org/wp-content/uploads/2018/02/Between-Two-Extremes-Feb-2018-ISD.pdf>

Anindya, C. R. (2019). The Deradicalisation Programme for Indonesian Deportees : A Vacuum in Coordination. *Journal for deradicalization*, 18, 217-243.

- Arce, D., et Sandler, T.** (2005). Counterterrorism : A Game-Theoretic Analysis. *The Journal of Conflict Resolution*, 49(2), 183-200. <https://doi.org/10.2307/30045107>
- Audit Commission.** (2008). *Preventing Violent Extremism : Learning and Development Exercise. Report to the Home Office and Communities and Local Government.* Audit Commission.
- Awan, I.** (2012a). Muslim communities, conflict and terrorism : A study of Alum Rock. *Safer Communities*, 11(4), 195-204. <https://doi.org/10.1108/17578041211271463>
- Awan, I.** (2012b). 'I am a Muslim not an extremist' : How the prevent strategy has constructed a « suspect » community. *Politics and Policy*, 40(6), 1158-1185. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1111/j.1747-1346.2012.00397.x>
- Azam, Z., et Bareeha, F.** (2017). Mishal : A Case Study of a Deradicalization and Emancipation Program in SWAT Valley, Pakistan. *Journal for Deradicalization*, 11, 1-29.
- Badurdeen, F. A., et Goldsmith, D. P.** (2018). Initiatives and Perceptions to Counter Violent Extremism in the Coastal Region of Kenya. *Journal for Deradicalization*, 16, 70-102.
- Bala, A., et Deman, H.** (2017). *Bottom-Up Approach to Countering Violent Extremism in Tunisia Final Evaluation Report* (p. 28). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2018/07/Final_External_Evaluation_Report_on_Bottom-Up_Approaches_to_CVE_Project_-_SFCG_Tunisia.pdf
- Barkindo, A., & Bryans, S.** (2016). De-Radicalising Prisoners in Nigéria : Developing a basic prison based de-radicalisation programme. *Journal for Deradicalization*, 7, 1-25.
- Baruch, B., Ling, T., Warnes, R., et Hofman, J.** (2018). Evaluation in an emerging field : Developing a measurement framework for the field of counter-violent-extremism. *Evaluation*, 24(4), 475-495.
- Basse, Y. O.** (2018). *Final Evaluation Kallewa Manio : An Integrated Approach to Counter Violent Extremism in Diffa* (p. 51). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2019/05/Final_Evaluation-Kallewa_Manio-June_2018.pdf
- Bastug, M. F., et Evlek, U. K.** (2016). Individual Disengagement and Deradicalization Pilot Program in Turkey : Methods and Outcomes. *Journal for Deradicalization*, 8, 25-45.
- Bean, S., Hill, P., Sany, J., et Riveles, S.** (2011). *USAID/West Africa Peace through Development (PDEV) : Program Assessment Report* (p. 87). USAID and EnCompass. https://pdf.usaid.gov/pdf_docs/PDACR829.pdf
- Beider, H., et Briggs, R.** (2010). Promoting community cohesion and preventing violent extremism in higher and further education. *Institute of community cohesion.*—URL: http://www.cohesioninstitute.org.uk/live/images/cme_resources/public/documents/publications/promoting-community-cohesion.pdf http://safecampuscommunities.ac.uk/uploads/files/2013/05/promoting_community_cohesion.pdf
- Bellasio, J., Hofman, J., Ward, A., Nederveen, F., Knack, A., Meranto, A. S., et Hoorens, S.** (2018). *Counterterrorism evaluation : Taking stock and looking ahead.* RAND. www.rand.org/t/RR2628
- Bilali, R.** (2019). 'Voices for Peace' Impact Evaluation of a Radio Drama to Counteract Violent Extremism in the Sahel Region in Burkina Faso—Endline Report (AID-OAA-M-13-00013 ; p. 116). USAID and NORC at the University of Chicago. https://pdf.usaid.gov/pdf_docs/pa00w4g3.pdf
- Bilazarian, T.** (2016). *Countering Violent Extremism : Lessons on Early Intervention from the United Kingdom's Channel Program* (p. 11). The Program on Extremism at George Washington University. <https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/downloads/Channel%20CVE%20UK.pdf>
- Björge, T., et Horgan, J.** (2009). *Leaving terrorism behind : Individual and collective disengagement* (37177830 ; 3892825). Routledge ; ProQuest Sociology Collection. <http://search.proquest.com/docview/37177830?accountid=28979>
- Boucek, C.** (2008). Counter-Terrorism from Within : Assessing Saudi Arabia's Religious Rehabilitation and Disengagement Programme. *The RUSI Journal*, 153(6), 60-65. <https://doi.org/10.1080/03071840802670106>
- Boucek, C.** (2009). Extremist re-education and rehabilitation in Saudi Arabia. In T. Björge et J. Horgan (Éds.), *Leaving Terrorism Behind* (p. 230-241). Routledge. <https://doi.org/10.4324/9780203884751-22>
- Boulton, A.** (2010). Education for development, CD for Peace : Producing the "globally competitive" child. *Geoforum*, 41(2), 329-336. <https://doi.org/10.1016/j.geoforum.2009.09.013>

- Bou Zeid, M. R.** (2019). Countering extremism through service learning : Narratives learned from journalism students. *Journal of Applied Research in Higher Education*, 11(4), 686-697. <https://doi.org/10.1108/JARHE-11-2018-0246>
- Bowie, R., et Revell, L.** (2018). How Christian Universities Respond to Extremism. *Education Sciences*, 8(3), 140. <https://doi.org/10.3390/educsci8030140>
- Boyd-MacMillan, E.** (2016). Increasing Cognitive Complexity and Collaboration Across Communities : Being Muslim Being Scottish. *Journal of Strategic Security*, 9(4), 79-110. <https://doi.org/10.5038/1944-0472.9.4.1563>
- Boyle, P., Bouasla, E., et Abderebbi, M.** (2016). *Mid-Term Evaluation Favorable Opportunities to Reinforce Self-Advancement for Today's Youth (FORSATY)* (AID-608-M-15-00002 ; p. 126). USAID.
- Brett, J., et Kahlmeyer, A.** (2017). *Strengthening Resilience to Violent Extremism – STRIVE (Horn of Africa) Evaluation Report* (p. 51). TANA, Conflict Management Consulting, The European Union. <https://ct-morse.eu/wp-content/uploads/2017/04/170124-STRIVE-evaluation-Report-Final.pdf>
- Briggs, R.** (2010). Community engagement for counterterrorism : Lessons from the United Kingdom. *International Affairs [London]*, 86(4), 971-981. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1111/j.1468-2346.2010.00923.x>
- Broadbent, R.** (2013). Using Grass Roots Community Programs as an Anti-Extremism Strategy. *Australian Journal of Adult Learning*, 53(2), 187-210. ERIC.
- Brooks, M. C., et Ezzani, M. D.** (2017). “Being Wholly Muslim and Wholly American” : Exploring One Islamic School’s Efforts to Educate Against Extremism. *Teachers College Record*, 119, 1-32.
- Brottsförebyggande rådet (Sweden) (Éd.).** (2001). *Exit : A follow-up and evaluation of the organisation for people wishing to leave racist and nazi groups*. Brottsförebyggande rådet. https://www.bra.se/download/18.cba82f7130f475a2f1800028108/1371914734840/2001_exit_a_follow-up_and_evaluation.pdf
- Bryan, H.** (2017). Developing the political citizen : How teachers are navigating the statutory demands of the Counter-Terrorism and Security Act 205 and the Prevent Duty. *Education, Citizenship and Social Justice*, 12(3), 213-226. <https://doi.org/10.1177/1746197917717841>
- Busher, J., Choudhury, T., et Thomas, P.** (2019). The enactment of the counter-terrorism “Prevent duty” in British schools and colleges : Beyond reluctant accommodation or straightforward policy acceptance. *Critical Studies on Terrorism*, 12(3), 440-462. <https://doi.org/10.1080/17539153.2019.1568853>
- Busher, J., Choudhury, T., Thomas, P., et Harris, G.** (2017). *What the Prevent duty means for schools and colleges in England : An analysis of educationalists’ experiences* (p. 68). Centre for Trust, Peace and Social Relations, Coventry University, Durham University, University of Huddersfield. <https://pdfs.semanticscholar.org/8371/af62c44e0eb9e0f35a3a24e9767d182f7299.pdf?ga=2.124265677.614667292.1610595013-1125023608.1610595013>
- Carthy, S. L., Doody, C. B., Cox, K., O’Hora, D., et Sarma, K. M.** (2020). Counter-narratives for the prevention of violent radicalisation : A systematic review of targeted interventions. *Campbell Systematic Reviews*, 16(3). <https://doi.org/10.1002/cl2.1106>
- Chatellier, S.** (2012). *Pakistani Women Moderating Extremism A Coalition-Building Case Study* (p. 18). The Institute for Inclusive Security. <https://www.inclusivesecurity.org/wp-content/uploads/2012/09/Pakistani-Women-Moderating-Extremism-A-Coalition-Building-Case-Study.pdf>
- Cherney, A.** (2020). Evaluating interventions to disengage extremist offenders : A study of the proactive integrated support model (PRISM). *Behavioral Sciences of Terrorism and Political Aggression*, 12(1), 17-36. <https://doi.org/10.1080/19434472.2018.1495661>
- Cherney, A., et Belton, E.** (2019). Evaluating Case-Managed Approaches to Counter Radicalization and Violent Extremism : An Example of the Proactive Integrated Support Model (PRISM) Intervention. *Studies in Conflict et Terrorism*, 1-21. <https://doi.org/10.1080/1057610X.2019.1577016>
- Chowdhury Fink, N., Romaniuk, P., et Barakat, R.** (2013). *Evaluating countering violent extremism programming : Practice and progress*. Center on Global Counterterrorism Cooperation.
- Christiaens, E., Hardyns, W., et Pauwels, L.** (2018). *Evaluating the BOUNCEUp Tool : Research Findings and Policy Implications* (p. 95). Ghent University, Faculty Law and Criminology, Institute for International Research on Criminal Policy (IRCP), Federal Public Service Home Affairs.
- Christmann, K., Rogerson, M., Hirschfield, A., Wilcox, A., et Sharratt K.** (2012). *Process Evaluation of Preventing Violent Extremism : Programmes for Young People*. <http://rgdoi.net/10.13140/21.3117.9042>

Cifuentes, R., Whittaker, G. R., et Lake, L. (2013). The Think Project : An Approach to Addressing Racism and Far-Right Extremism in Swansea, South Wales. *Democracy and Security*, 9(3), 304-325. Sociological Abstracts. <https://doi.org/10.1080/17419166.2013.802985>

Cipaku, J. (2013). *Mid-Term Evaluation Reducing Recidivism : A Process for Effective Disengagement of High-Risk Prisoners in Indonesia* (N° 7 ; p. 22). Search for Common Ground and New Zealand's International Aid et Development Agency. https://www.sfcg.org/wp-content/uploads/2014/07/INA_MT_Dec13_SCGF_NZL_MTR_Report_Revisi.pdf

Clemens-Hope, O. M. (2015). *USAID Peace Through Development II* (p. 56). USAID. https://blumont.org/wp-content/uploads/2016/01/PDev-II-Jan-Mar-2015-Qly-Rept_FINAL-4-29-2015.pdf

Clement, P.-A., Madriaza, P., et Morin, D. (2021). *Les intervenants et l'évaluation en prévention de l'extrémisme violent : Entre contraintes et opportunités*. Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (Chaire UNESCO-PREV).

Cockayne, J., O'Neil, S., Felbab-Brown, V., Chowdhury Fink, N., et Oswald, B. "Ossie". (2015). *UN DDR in an Era of Violent Extremism : Is It Fit for Purpose ?* (p. 164). United Nations University. https://collections.unu.edu/eserv/UNU:5532/UN_DDR_in_an_Era_of_Violent_Extremism_2018.pdf

Colibaba, A., Colibaba, S., Gheorghiu, I., Colibaba, C., Dinu, C., et Ursa, O. (2017). The Xeno-Tolerance Project- A Useful Tool in Doing Quality Research. *Lucrări Științifice*, 60(1), 141-144.

Connell, J. P., et Kubisch, A. C. (1998). Applying a theory of change approach to the evaluation of comprehensive community initiatives : Progress, prospects, and problems. *New approaches to evaluating community initiatives*, 2(15-44), 1-16.

Court, D. (2006). Foolish Dreams in a Fabled Land : Living Co-Existence in an Israeli Arab School. *Curriculum Inquiry*, 36(2), 189-208. <https://doi.org/10.1111/j.1467-873X.2006.00352.x>

CPN-PREV. (2020). *Train the Trainers Manual : A toolkit to facilitate training on the prevention of violent radicalization for practitioners across Canada*. Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).

Cragin, K., et Chalk, P. (2003). *Terrorism & development : Using social and economic development to inhibit a resurgence of terrorism* (1st éd.). Rand. https://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR1630.pdf

Davey, J., Birdwell, J., et Skellett, R. (2018). *Counter Conversations A model for direct engagement with individuals showing signs of radicalisation online* (p. 32). Institute for Strategic Dialogue (ISD). https://www.isdglobal.org/wp-content/uploads/2018/03/Counter-Conversations_FINAL.pdf

Davey, J., Tuck, H., et Amarasingam, A. (2019). *An imprecise science : Assessing interventions for the prevention, disengagement and de-radicalisation of left and right-wing extremists*. Institute for Strategic Dialogue.

Demant, F., Wagenaar, W., et van Donselaar, J. (2009). *Racism & Extremism Monitor Deradicalisation in practice*. Amsterdam, Netherlands: Leiden University, Anne Frank House. http://www.annefrankdagboek.nl/ImageVaultFiles/id_12097/cf_21/Deradicalisation_ebook.PDF

Dhungana, S. K., Ismanbaeva, R., et Aisakhunova, A. (2016). *Reducing Violent Religious Extremism and Preventing Conflict in Kyrgyzstan and Central Asia 2013-2016* (p. 69). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2017/04/CSSF_KGZ501-Review-Final-Report_13052016.pdf

Dietrich, K. (2018). *The Way Forward, Assessing the Impact of the « White Dove » CVE Radio Project in Northern Nigéria* (p. 33). Equal Access International. <https://www.equalaccess.org/wp-content/uploads/2018/11/2018-EAI-Nigéria-White-Dove-Final-Assessment.pdf>

Dunn, K. M., Atie, R., Kennedy, M., Ali, J. A., O'Reilly, J., et Rogerson, L. (2015). Can you use community policing for counter terrorism ? Evidence from NSW, Australia. *Police Practice and Research*, 17(3), 196-211. <https://doi.org/10.1080/15614263.2015.1015126>

Dwyer, C. D. (2010). « *Sometimes I wish I was an'ex'ex-prisoner* » : *Release and reintegration : The experience of politically motivated former prisoners in Northern Ireland* [PhD Thesis, Queen's University]. https://pureadmin.qub.ac.uk/ws/portalfiles/portal/178891096/Dwyer_Sometimes_I_69438803.pdf

Dwyer, C. D., et Maruna, S. (2011). The Role of Self-Help Efforts in the Reintegration of 'Politically Motivated' Former Prisoners : Implications from the Northern Irish Experience. *Crime, Law and Social Change*, 55(4), 293-309. <https://doi.org/10.1007/s10611-011-9284-7>

Education Development Center (EDC) et USAID. (2019). *USAID's Mindanao Youth for Development (MYDev) Program FY17 Impact Evaluation Report & FY18/19 (Extension) Performance Evaluation Report Measuring Youth's Employment, Perceptions and Engagements, and Skills* (p. 33). USAID and Education Development Center (EDC).

El-Said, H. (2015). *New approaches to countering terrorism : Designing and evaluating counter radicalization and de-radicalization programs*. <http://www.palgraveconnect.com/doi/10.1057/9781137449979>

Elwick, A., et Jerome, L. (2019). Balancing securitisation and education in schools : Teachers' agency in implementing the Prevent duty. *Journal of Beliefs & Values*, 40(3), 338-353. <https://doi.org/10.1080/13617672.2019.1600322>

Eriksson, A. (2008). Challenging cultures of violence through community restorative justice in Northern Ireland. In *Sociology of Crime Law and Deviance* (Vol. 11, p. 231-260). Emerald (MCB UP). [https://www.emerald.com/insight/content/doi/10.1016/S1521-6136\(08\)00410-7/full/html](https://www.emerald.com/insight/content/doi/10.1016/S1521-6136(08)00410-7/full/html)

Feddes, A. R. (2015). *Socio-psychological factors involved in measures of disengagement and deradicalization and evaluation challenges in Western Europe*. <http://www.mei.edu/sites/default/files/Feddes.pdf>

Feddes, A. R., et Gallucci, M. (2015). A literature review on methodology used in evaluating effects of preventive and de-radicalisation interventions. *Journal for Deradicalization*, 5, 1-27.

Feddes, A. R., Huijzer, A., van Ooijen, I., et Doosje, B. (2019). Fortress of Democracy : Engaging Youngsters in Democracy Results in More Support for the Political System. *Peace and Conflict: Journal of Peace Psychology*, 25(2), 158-164. <https://doi.org/10.1037/pac0000380>

Feddes, A. R., Mann, L., et Doosje, B. (2015). Increasing self-esteem and empathy to prevent violent radicalization : A longitudinal quantitative evaluation of a resilience training focused on adolescents with a dual identity: Increasing self-esteem and empathy to prevent violent radicalization. *Journal of Applied Social Psychology*, 45(7), 400-411. <https://doi.org/10.1111/jasp.12307>

Finkel, S. E., Belasco, C. A., Gineste, C., Neureiter, M., et McCauley, J. (2018). *Peace Through Development II Burkina Faso, Chad, and Niger Impact Evaluation Endline Report* (p. 149). USAID. https://pdf.usaid.gov/pdf_docs/PA00SWPK.pdf

Finkel, S. E., Belasco, C. A., Neureiter, M., McCauley, J., Hoepers, B., et Corrigan, C. C. (2017). *USAID/WEST Africa (USAID/WA) Evaluation & Analytical Services (EAS) Project for the Regional Peace and Governance Programs, Midline Report for Impact Evaluation of the Peace Through Development Phase II (PDEV II) Project in Chad, Niger, and Burkina Faso* (p. 163). USAID.

Finkel, S. E., Rojo-Mendoza, R. T., Schwartz, C. L., Belasco, C. A., et Kreft, A. (2015). *Evaluation & Analytical Services (EAS) Project for The Regional Peace and Governance Programs Impact Evaluation of Peace through Development II (P-DEV II) Radio Programming in Chad and Niger Final Report* (p. 50). USAID and University of Pittsburgh. https://pdf.usaid.gov/pdf_docs/PA00KTF3.pdf

Finn, M., Momani, B., Opatowski, M., et Opondo, M. (2016). Youth Evaluations of CVE/PVE Programming in Kenya in Context. *Journal for Deradicalization*, 7, 164-224.

Franssen, A., Dal, C., et Rinschbergh, F. (2019). *Rapport d'évaluation du Réseau de prise en charge des radicalismes et extrémismes violents* (p. 186). Centre d'Études Sociologiques (CES) de l'Université Saint-Louis - Bruxelles (USL-B). https://extremismes-violents.cfwb.be/fileadmin/sites/RAR/uploads/Documents_Reseau/Rapport_final_Evaluation_Reseau-FWB_11-11-2019.pdf

Frenett, R., et Dow, M. (2015). *One to One Online Interventions A pilot CVE methodology* (p. 28). Institute for Strategic Dialogue (ISD). https://www.isdglobal.org/wp-content/uploads/2016/04/One2One_Web_v9.pdf

Garaigordobil, M. (2012). Evaluation of a program to prevent political violence in the Basque conflict : Effects on the capacity of empathy, anger management and the definition of peace. *Gaceta Sanitaria*, 26(3), 211-216. <https://doi.org/10.1016/j.gaceta.2011.06.014>

Gatewood, C., et Boyer, I. (2019). *Building Digital Citizenship in France Lessons from the Sens Critique project* (p. 36). Institute for Strategic Dialogue (ISD). https://www.isdglobal.org/wp-content/uploads/2019/03/Sens-Critique-uk-screen_E3.pdf

Gielen, A.-J. (2017). Countering Violent Extremism : A Realist Review for Assessing What Works, for Whom, in What Circumstances, and How ? *Terrorism and Political Violence*, 31(6), 1149-1167. <https://doi.org/10.1080/09546553.2017.1313736>

Gill, P., Clemmow, C., Hetzel, F., Rottweiler, B., Salman, N., Van Der Vegt, I., Marchment, Z., Schumann, S., Zolghadriha, S., et Schulten, N. (2020). Systematic Review of Mental Health Problems and Violent Extremism. *The Journal of Forensic Psychiatry & Psychology*, 1-28.

Glazzard, A., et Reed, A. (2018). *Global Evaluation of the European Union Engagement on Counter-Terrorism* (p. 35). Royal United Services Institute and International Center for Counter-Terrorism - The Hague (ICCT). <https://icct.nl/app/uploads/2018/10/eu-ct-evaluation-v7-final.pdf>

Goaziou, V. L. (2018). L'éducatif au prisme de la radicalisation—La Cellule d'écoute et d'accompagnement des familles (CEAF) de l'ADDAP13. *Recherches et pratiques pour le Groupe addap13*, 2, 52.

Government of the United Kingdom. (2011). *Prevent Strategy : Equality Impact Assessment* (p. 17). Government of the United Kingdom. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/97979/prevent-review-eia.pdf

Greiner, K. (2010). *Applying Local Solutions to Local Problems : Radio Listeners as Agents of Change* (IQC DFD-I-00-05-00244-00 ; p. 57). USAID and University of South Florida. https://pdf.usaid.gov/pdf_docs/PDACS538.pdf

Harahap, H. I., Irmayani, T., et Lubis, F. H. (2019). The Rationality of De-radicalization Efforts for the Children of Terrorists at Al-Hidayah Islamic Boarding School. *International Journal of Islamic Thought*, 16(1), 38-50. <https://doi.org/10.24035/ijit.16.2019.004>

Harris-Hogan, S., Barrelle, K., et Smith, D. (2019). The role of schools and education in countering violent extremism (CVE) : Applying lessons from Western countries to Australian CVE policy. *Oxford Review of Education*, 45(6), 731-748. <https://doi.org/10.1080/03054985.2019.1612343>

Hassan, G., Brouillette-Alarie, S., Ousman, S., Kilinc, D., Varela, W., Lavoie, L., Fetiu, A., Harris-Hogan, S., Madriaza, P., Borokhovski, E., Pickup, D., Boivin, M., Srimathi Narayana, M., Rousseau, C., Gill, P., Thompson, S., McCoy, J., Venkatesh, V., et Morin, D. (2021). *A Systematic Review on the Outcomes of Primary and Secondary Prevention Programs in the Field of Violent Radicalization*. Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV).

Hassan, G., Brouillette-Alarie, S., Ousman, S., Savard, É., Kilinc, D., Madriaza, P., Varela, W., Pickup, D., Danis, E., et CPN-PREV team. (2021). *A Systematic Review on the Outcomes of Tertiary Prevention Programs in the Field of Violent Radicalization*. Canadian Practitioners Network for the Prevention of Radicalization and Extremist Violence (CPN-PREV). <https://cpnprev.ca/wp-content/uploads/2021/11/Intervention-report-FINAL-2021.pdf>

Hassan, G., Ousman, S., Madriaza, P., Fetiu, A., Boily, L.-A., Levesque, F., Squalli, Z., Ajrouche, K., El-Tahry, N., Lampron-De Souza, S., Desmarais, L., Duong, E., et Moyano, R. (2020). *D'un océan à l'autre : Cartographie des initiatives de prévention secondaire et tertiaire oeuvrant dans un contexte de radicalisation et d'extrémisme violent au Canada*. Réseau des Praticiens Canadiens pour la Prévention de la Radicalisation et de l'Extrémisme Violent (RPC-PREV). <https://cpnprev.ca/wp-content/uploads/2020/12/FR-Rapport-mapping-final-1.pdf>

Heath-Kelly, C., et Strausz, E. (2018). The banality of counterterrorism “after, after 9/11” ? Perspectives on the Prevent duty from the UK health care sector. *Critical Studies on Terrorism*, 12(1), 89-109. <https://doi.org/10.1080/17539153.2018.1494123>

Helmus, T., et Klein, K. (2018). *Assessing Outcomes of Online Campaigns Countering Violent Extremism : A Case Study of the Redirect Method* (p. 19). RAND Corporation. https://www.rand.org/pubs/research_reports/RR2813.html

Heydemann, S. (2014). Countering violent extremism as a field of practice. *Insights*, 1-4.

Hiariej, E., Rachmawati, A. D., Taek, A. M., Kurniasari, M., et Alvian, R. A. (2017). *Final Evaluation Reducing the Recruitment and Recidivism of Violent Extremists in Indonesia* (p. 122). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2018/01/INA029_DOS_BC_external_Evaluation_Report_FINAL_2017.pdf

Hirschi, C., et Widmer, T. (2012). Approaches and challenges in evaluating measures taken against right-wing extremism. *Evaluation and program planning*, 35(1), 171-179.

Holmer, G. (2013). Countering Violent Extremism : A Peacebuilding Perspective. *USIP Special Report*, 336. <http://www-preview.usip.org/sites/default/files/SR336-Countering%20Violent%20Extremism-A%20Peacebuilding%20Perspective.pdf>

Holmer, G., Bauman, P., et Aryaeinejad, K. (2018). *Measuring Up : Evaluating the Impact of P/CVE Programs* (p. 2018-09). United States Institute of Peace.

Hong, Q. N., & Pluye, P. (2019). A conceptual framework for critical appraisal in systematic mixed studies reviews. *Journal of Mixed Methods Research*, 13(4), 446-460.

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., et Nicolau, B. (2018). *Mixed Methods Appraisal Tool (MMAT) Version 2018—User guide*. McGill University. http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/attach/127916259/MMAT_2018_criteria%C3%A2%E2%82%AC%20manual_2018%C3%A2%E2%82%AC%2008%C3%A2%E2%82%AC%2001_ENG.pdf

Horgan, J., et Braddock, K. (2010). Rehabilitating the Terrorists ? : Challenges in Assessing the Effectiveness of De-radicalization Programs. *Terrorism and Political Violence*, 22(2), 267-291. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1080/09546551003594748>

Iacopini, G., Stock, L., et Junge, K. (2011). *Evaluation of Tower Hamlets Prevent Projects* (p. 63). The Tavistock Institute. https://www.tavistock.org/wp-content/uploads/2012/12/Tavistock_Projects_Tower-Hamlets-PVE_evaluation_final_report.pdf

Ipp, O., Prado, A., et Fourati, Y. A. (2014). *Tunisia Transition Initiative (TTI) Final Evaluation Report* (AID-OAA-I-13-00007 ; p. 140). USAID and The QED Group. https://pdf.usaid.gov/pdf_docs/PA00JZD7.pdf

Istiqomah, M. (2011). *De-radicalization program in Indonesian prisons : Reformation on the correctional institution*. 29-34. <https://doi.org/10.4225/75/57a4200e2b5a3>

i-works research ltd. (2013). *The Think Project Interim Evaluation* (p. 26). Ethnic Youth Support Team (EYST).

Jackson, B., Rhoades, A., Reimer, J., Lander, N., Costello, K., et Beaghley, S. (2019). *Practical Terrorism Prevention : Reexamining U.S. National Approaches to Addressing the Threat of Ideologically Motivated Violence*. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2647.html

Jailobaeva, K., et Asilbekova, G. (2017). *Social Media for Deradicalization in Kyrgyzstan : A model for Central Asia Final project evaluation* (p. 58). Search for Common Ground. <https://www.dmeformpeace.org/resource/final-evaluation-social-media-for-deradicalization-in-kyrgyzstan-a-model-for-central-asia/>

Jerome, L., et Elwick, A. (2016). *Evaluation report on the ACT Building Resilience Project Full Report* (p. 64). Middlesex University London and Association for Citizenship Teaching (ACT). <https://eprints.mdx.ac.uk/22857/1/The%20ACT%20Building%20Resilience%20Project%20Full%20Evaluation%20Report%20FINAL%2031.10.16.pdf>

Jerome, L., et Elwick, A. (2019). Identifying an Educational Response to the Prevent Policy : Student Perspectives on Learning about Terrorism, Extremism and Radicalisation. *British Journal of Educational Studies*, 67(1), 97-114. <https://doi.org/10.1080/00071005.2017.1415295>

Johns, A., Grossman, M., et McDonald, K. (2014). « More Than a Game » : The Impact of Sport-Based Youth Mentoring Schemes on Developing Resilience toward Violent Extremism. *Social Inclusion*, 2(2). ProQuest Sociology. <https://doi.org/10.17645/si.v2i2.167>

Johnston, D., Hussain, A., et Cataldi, R. (2008). *Madrassa Enhancement and Global Security A Model for Faith-Based Engagement*. <https://icrd.org/wp-content/uploads/2018/01/Madrassa-Enhancement-and-Global-Security.pdf>

Joyce, C. A. (2018). *Exploring teachers' beliefs, values and attitudes towards radicalisation, extremism and the implementation of anti-radicalisation strategies* [The University of Sheffield School of Education]. <https://etheses.whiterose.ac.uk/21452/1/Craig%20Joyce%20Thesis%20August%202018.pdf>

Khalil, J., Brown, R., Chant, C., Olowo, P., et Wood, N. (2019). *Deradicalisation and Disengagement in Somalia—Evidence from a Rehabilitation Programme for Former Members of Al-Shabaab* (p. 50). Royal United Services Institute for Defence and Security Studies. https://rusi.org/sites/default/files/20190104_whr_4-18_deradicalisation_and_disengagement_in_somalia_web.pdf

Khalil, J., et Ipp, O. (2016). *USAID/OTI PDQIII Task Order #10, Activity #3 Mali Transition Initiative : Final Evaluation* (Q011OAA1500012 ; p. 62). USAID.

Khalil, J., et Zeuthen, M. (2014). A Case Study of Counter Violent Extremism (CVE) Programming : Lessons from OTI's Kenya Transition Initiative. *Stability: International Journal of Security & Development*, 3(1), 12. <https://doi.org/10.5334/sta.ee>

- Khurshid, K., Bhatti, A. J., et Hussain, B.** (2018). Education for Social Justice : Commitments and Achievements. *Pakistan Journal of Social Sciences*, 38(1), 199-219.
- Knox, C., & Hughes, J.** (1996). Crossing the Divide : Community Relations in Northern Ireland. *Journal of Peace Research*, 33(1), 83-98.
- Kollmorgen, J.-C., et Barry, C.** (2017). *Evaluation Report Ex-Post Performance Evaluation of USAID/RDMA Sapan Program* (AID-486-I-14-00001/ AID-486-TO-16-00007 ; p. 116). USAID. <https://pdf.usaid.gov/pdf/docs/PA00MXJ8.pdf>
- Kollmorgen, J.-C., Ogada, M., Korir, S., et Dena, E.** (2019). *Strengthening Community Resilience Against Extremism (SCORE) Mid-Term Performance Evaluation—Final Report* (AID-OAA-M-13-00011 ; p. 114). USAID and Social Impact, Inc.
- Kruglanski, A. W., Gelfand, M. J., Bélanger, J. J., Sheveland, A., Hetiarachchi, M., et Gunaratna, R.** (2014). The Psychology of Radicalization and Deradicalization : How Significance Quest Impacts Violent Extremism: Processes of Radicalization and Deradicalization. *Political Psychology*, 35, 69-93. <https://doi.org/10.1111/pops.12163>
- Kundnani, A.** (2009). *Spooked ! How not to prevent violent extremism*. Institute of Race Relations. <https://www.kundnani.org/wp-content/uploads/spooked.pdf>
- Kundnani, A.** (2012). Radicalisation : The journey of a concept. *Race & Class*, 54(2), 3-25. International Bibliography of the Social Sciences (IBSS) ; Sociological Abstracts. <https://doi.org/10.1177/0306396812454984>
- Kurtz, J.** (2015). *Does Youth Employment Build Stability ? Evidence from an Impact Evaluation of Vocational Training in Afghanistan* (p. 40). Mercy Corps. https://www.mercycorps.org/sites/default/files/2020-01/MercyCorps_AfghanistanINVEST_ImpactEvaluation_2015.pdf
- Kyriacou, C., Szczepek Reed, B., Said, F., et Davies, I.** (2017). British Muslim university students' perceptions of Prevent and its impact on their sense of identity. *Education, Citizenship and Social Justice*, 12(2), 97-110. <https://doi.org/10.1177/1746197916688918>
- Lakhani, S.** (2012). Preventing violent extremism : Perceptions of policy from grassroots and communities. *Howard Journal of Criminal Justice*, 51(2), 190-206. ProQuest Sociology Collection. <https://doi.org/10.1111/j.1468-2311.2011.00685.x>
- Lamhaidi, N.** (2017). *Women's Caravan for Peace Final Evaluation* (p. 18). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2018/07/WC4P-Evaluation_EN-FINAL.pdf
- Letsch, L.** (2018). Countering Violent Extremism in Tunisia – Between Dependency and Self-Reliance. *Journal for Deradicalization*, 17, 163-195.
- Levy, D., Jamankulov, K., et Sartbay, T.** (2019). *Project Evaluation #JashStan : Youth as Agents of Peace and Stability in Kyrgyzstan* (p. 73). Evidence Research Institute. <https://www.dme4peace.org/resource/final-evaluation-jashstan-youth-as-agents-of-peace-and-stability-in-kyrgyzstan/>
- Liht, J., et Savage, S.** (2013). Preventing Violent Extremism through Value Complexity : Being Muslim Being British. *Journal of Strategic Security*, 6(4), 44-66. <https://doi.org/10.5038/1944-0472.6.4.3>
- Lindekilde, L.** (2012a). Neo-liberal Governing of « Radicals » : Danish Radicalization Prevention Policies and Potential Iatrogenic Effects. *International Journal of Conflict and Violence*, 6(1), 109-125. ProQuest Sociology Collection.
- Lindekilde, L.** (2012b). Value for Money ? Problems of Impact Assessment of Counter-Radicalisation Policies on End Target Groups : The Case of Denmark. *European Journal on Criminal Policy and Research*, 18(4), 385-402. ProQuest Central ; ProQuest Sociology. <https://doi.org/10.1007/s10610-012-9178-y>
- Lindekilde, L.** (2014). Refocusing Danish counter-radicalisation efforts : An analysis of the (problematic) logic and practice of individual de-radicalisation interventions. In *Counter-Radicalisation* (p. 223-241). Routledge. <https://doi.org/10.4324/9781315773094-14>
- Lobnikar, B., Cajner Mraović, I., et Prislan, K.** (2019). Preventing Radicalisation and Extremism – The Views of Police Students in Croatia. *VARSTVOSLOVJE*, 2, 161-183.
- Lum, C., Kennedy, L. W., et Sherley, A. J.** (2006). The Effectiveness of Counter-Terrorism Strategies : Campbell Systematic Review Summary. *Campbell systematic reviews*, 2(1), 1-50.
- Mackenzie, M., et Blamey, A.** (2005). The Practice and the Theory : Lessons from the Application of a Theories of Change Approach. *Evaluation*, 11(2), 151-168. <https://doi.org/10.1177/1356389005055538>

Madriaza, P., Morin, D., Ousman, S., Autixier, C., Hassan, G., et Venkatesh, V. (2021). *Améliorer l'évaluation en prévention de l'extrémisme violent, c'est l'affaire de tous ! Une étude exploratoire internationale*. Chaire UNESCO en prévention de la radicalisation et de l'extrémisme violents (Chaire UNESCO-PREV).

Madriaza, P., et Ponsot, A.-S. (2015). *Preventing radicalization : A systematic review*. International Centre for the Prevention of Crime.

Madriaza, P., Ponsot, A.-S., Marion, D., Monnier, C., Ghanem, A., Zaim, B., Nait-Kac, W., Hassani, N., et Autixier, C. (2017). *The prevention of radicalization leading to violence : An international study of front-line workers and intervention issues* (p. 137). International Centre for the Prevention of Crime. http://www.crime-prevention-intl.org/fileadmin/user_upload/Publications/2017/EN_Rapport_Radicalisation_Final_Aout2017.pdf

Madriaza, P., Valendru, F., Stock-Rabbat, L., Ponsot, A.-S., et Marion, D. (2018). *Dispositif d'intervention sur la radicalisation violente en milieu ouvert (SPIP) en France* (p. 146). International Centre for the Prevention of Crime.

Manby, M. (2009a). *Evaluation of Kirklees Youth Offending Team (PVE) Pilot Parenting Project*. Ravensthorpe Community Centre (p. 25). Nationwide Children's Research Centre.

Manby, M. (2009b). *Kirklees Youth Offending Team Prevent (previously PVE) Programme (3) Evaluation of Theatre Project* (p. 21). Nationwide Children's Research Centre.

Manby, M. (2009c). *Kirklees Youth Offending Team. Prevent (previously PVE) Programme—(2) Evaluation of Film Project* (p. 41). Nationwide Children's Research Centre.

Manby, M. (2009d). *Kirklees Youth Offending Team. Prevent (previously PVE) Programme—Evaluation of Diversity Group* (p. 25). Nationwide Children's Research Centre.

Manby, M. (2010a). *Evaluation of Kirklees Youth Offending Team Prevent (previously PVE) Project Evaluation of Citizenship Programme* (p. 25). Nationwide Children's Research Centre.

Manby, M. (2010b). *Kirklees Youth Offending Team Prevent (previously PVE) Project Evaluation of Pathways into Adulthood Programme* (p. 28). Nationwide Children's Research Centre.

Mansour, S. (2017). *The Morocco Transforming Violent Extremism Media Training Program Final External Evaluation* (p. 57). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2018/06/MAR039_Final_Evaluation.pdf

Marret, J.-L., Bellasio, J., van den Berg, H., van Gorp, A., van Hemert, D., Leone, L., Meijer, R., Warnes, R., et Van Wonderen, R. (2017). *Final report providing background for using and further developing the validated toolkit Impact* (FP7 GA no. 312235). Impact Europe.

Mastroe, C. (2016). Evaluating CVE: Understanding the Recent Changes to the United Kingdom's Implementation of Prevent. *Perspectives on Terrorism*, 10(2), 50-60.

Mastroe, C., et Szmania, S. (2016). *Surveying CVE Metrics in Prevention, Disengagement and Deradicalization Programs*. Report to the Office of University Programs, Science and Technology Directorate, Department of Homeland Security.

McDonald, B., et Mir, Y. (2011). Al-Qaida-influenced violent extremism, UK government prevention policy and community engagement. *Journal of Aggression, Conflict and Peace Research*, 3(1), 32-44. ProQuest Sociology.

McDowell-Smith, A., Speckhard, A., et Yayla, A. S. (2017). Beating ISIS in the Digital Space : Focus Testing ISIS Defector Counter-Narrative Videos with American College Students. *Journal for Deradicalization*, 10, 50-76.

McGlynn, C., et McDaid, S. (2016). Radicalisation and Higher Education : Students' Understanding and Experiences. *Terrorism and Political Violence*, 31(3), 559-576. <https://doi.org/10.1080/09546553.2016.1258637>

McRae, D. (2009a). DDR and Localized Violent Conflict : Evaluating Combatant Reintegration Programs in Poso, Indonesia. *Indonesian Social Development Paper*, 14.

McRae, D. (2009b). *Reintegration and localized conflict : Promoting police-combatant communication*. The World Bank.

McRae, D. (2010). Reintegration and localised conflict : Security impacts beyond influencing spoilers. *Conflict, Security & Development*, 10(3), 403-430. <https://doi.org/10.1080/14678802.2010.484204>

- Meringolo, P., Bosco, N., Cecchini, C., et Guidi, E.** (2019). Preventing violent radicalization in Italy : The actions of EU project PROVA. *Peace and Conflict: Journal of Peace Psychology*, 25(2), 165-169. <https://doi.org/10.1037/pac0000375>
- Mitts, T.** (2017). Do Community Engagement Efforts Reduce Extremist Rhetoric on Social Media ? *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2940290>
- Moffett, K., et Sgro, T.** (2016). School-Based CVE Strategies. *The ANNALS of the American Academy of Political and Social Science*, 668(1), 145-164. <https://doi.org/10.1177/0002716216672435>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., et The PRISMA Group.** (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses : The PRISMA Statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Monzani, B., Sarota, A., et Venturi, B.** (2018). *Evaluation Report Inuka ! Community-Led Security Approaches to Violent Extremism in Coastal Kenya* (p. 35). Search for Common Ground and Agency for Peacebuilding. <https://www.dmeformpeace.org/resource/final-evaluation-inuka-community-led-security-approaches-to-violent-extremism-in-coastal-kenya-october-2018/>
- Muncy, D., David, R., et Saleh, B.** (2015). *Training of Leaders on Religious and National Co-Existence (TOLERANCE) Project Mid-term Evaluation Report* (p. 113). USAID. https://pdf.usaid.gov/pdf_docs/PA00KMKD.pdf
- Murtaza, N., Sohail, A., Perveen Shaikh, R., Ahmed, S., Anver, S., Ahmad, A., Asghar, M., Ashraf, J., et Yar Khan, U.** (2018). *Punjab Youth Workforce Development Project Midterm Evaluation* (AID-391-C-15-00004 ; p. 116). USAID.
- Neumann, P.** (2008). Introduction. In J. ICST (Éd.), *Perspectives on Radicalisation and Political Violence : Papers from the First International Conference on Radicalisation and Political Violence*. International Centre for the Study of Radicalisation and Political Violence. <https://www.nonviolent-conflict.org/wp-content/uploads/2016/11/Perspectives-on-Radicalisation-Political-Violence.pdf>
- Neumann, P.** (2011). *Preventing violent radicalization in America*. Bipartisan Policy Center.
- Neumann, P., et Kleinmann, S.** (2013). How rigorous is radicalization research ? *Democracy and Security*, 9(4), 360-382. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1080/17419166.2013.802984>
- Nicolls, M., et Hassan, A.** (2014). *Evaluation Report Mid-Term Performance Evaluation of the USAID Somali Youth Leaders Initiative (SYLI)* (AID-RAN-I-00-09-00016 ; p. 98). USAID. https://pdf.usaid.gov/pdf_docs/PA00K3XD.pdf
- Octavia, L., et Wahyuni, E.** (2014). *Final Evaluation Report For the Project : Countering & Preventing Radicalization in Indonesian Pesantren* (p. 75). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2014/08/DUT_Evaluation_Report_FINAL.pdf
- Onyima, J. K.** (2017). Sub-Saharan Africa : Societal Reintegration of Ex-Militant Youths. *Conflict Studies Quarterly*, 21, 76-100. <https://doi.org/10.24193/cs.q.21.4>
- Orban, F.** (2019). Le programme de mentorat norvégien pour détenus radicalisés : Premier bilan, premiers enseignements « La prison au-delà des frontières ». *Les presses de l'ENAP*, 13.
- O'Toole, T., DeHanas, D. N., et Modood, T.** (2012). Balancing tolerance, security and Muslim engagement in the United Kingdom : The impact of the 'Prevent' agenda. *Critical Studies on Terrorism*, 5(3), 373-389. <https://doi.org/10.1080/17539153.2012.725570>
- O'Toole, T., Meer, N., DeHanas, D. N., Jones, S. H., et Modood, T.** (2016). Governing through prevent ? Regulation and contested practice in State-Muslim engagement. *Sociology*, 50(1), 160-177.
- O'Toole, T., DeHanas, D., Modood, T., Meer, N., et Jones, S.** (2013). *Taking Part Muslim Participation in Contemporary Governance* (p. 72). Centre for the Study of Ethnicity and Citizenship. <http://www.bristol.ac.uk/media-library/sites/ethnicity/migrated/documents/mpcgreport.pdf>
- Parker, D., et Lindekilde, L.** (2020). Preventing Extremism with Extremists : A Double-Edged Sword ? An Analysis of the Impact of Using Former Extremists in Danish Schools. *Education Sciences*, 10(4), 1-19. <https://doi.org/10.3390/educsci10040111>
- Parker, L., Boyer, I., et Gatewood, C.** (2018). *Young Digital Leaders Impact Report* (p. 44). Institute for Strategic Dialogue (ISD). https://www.isdglobal.org/wp-content/uploads/2018/10/YDL_Impact-Report_Final_October_2018.pdf

- Peracha, F. N., Khan, R. R., et Savage, S.** (2016). Sabaoon : Educational methods successfully countering and preventing violent extremism. In *Expanding Research on CVE* (p. 84-104).
- Peterson, A.** (2012). Legitimacy and the Swedish Security Service's Attempts to Mobilize Muslim Communities. *International journal of criminology and sociology*, 1, 109-120.
- Piasecka, S.** (2019). Performing PREVENT : Anti-extremist theatre-in-education in the service of UK counter-terrorism, a Freirean analysis. *Critical Studies on Terrorism*, 12(4), 715-734. <https://doi.org/10.1080/017539153.2019.1615660>
- Pickering, S., McCulloch, J., et Wright-Neville, D.** (2008). Counter-terrorism policing : Towards social cohesion. *Crime, Law and Social Change*, 50(1-2), 91-109. <https://doi.org/10.1007/s10611-008-9119-3>
- Pipe, R., Egal, J., Malla, L., Billow, Z., et Abdi, A.** (2016). *Somalia Program Support Services Final Performance Evaluation of the Transition Initiatives for Stabilization Project* (IDIQ AID-623-I-14-00009 ; p. 156). USAID and International Business & Technical Consultants, Inc. (IBTCI). <https://static1.squarespace.com/static/5db70e83fc0a966cf4cc42ea/t/5f491ed3020a2654cb8d19b7/1598627541959/1344.pdf>
- Pistone, I., Eriksson, E., Beckman, U., Mattson, C., et Sager, M.** (2019). A scoping review of interventions for preventing and countering violent extremism : Current status and implications for future research. *Journal for Deradicalization*, 19, 1-84.
- Powers, S. T.** (2015). Expanding the Paradigm : Countering Violent Extremism in Britain and the Need for a Youth Centric Community Based Approach. *Journal of Terrorism Research*, 6(1), 19-26. <https://doi.org/10.15664/jtr.1074>
- Pratchett, L., Thorp, L., Wingfield, M., Lowndes, V., et Jabbar, R.** (2010). *Preventing Support for Violent Extremism through Community Interventions : A Review of the Evidence-Rapid Evidence Assessment Full Final Report*. Department for Communities and Local Government.
- Ranstorp, M.** (2010). *Understanding violent radicalisation : Terrorist and jihadist movements in Europe* (889171749 ; 4230150). Routledge ; International Bibliography of the Social Sciences (IBSS). <http://search.proquest.com/docview/889171749?accountid=28004>
- Reeves, J., et Crowther, T.** (2019). Teacher feedback on the use of innovative social media simulations to enhance critical thinking in young people on radicalisation, extremism, sexual exploitation and grooming. *Pastoral Care in Education*, 37(4), 280-296. <https://doi.org/10.1080/02643944.2019.1618377>
- Reichardt, C. S.** (2009). Quasi-Experimental Design. In *The SAGE Handbook of Quantitative Methods in Psychology* (p. 47-72). SAGE Publications Ltd. <https://doi.org/10.4135/9780857020994>
- Reynolds, L.** (2017). *Internet Citizens Impact Report* (p. 50). Institute for Strategic Dialogue (ISD). <https://www.isdglobal.org/wp-content/uploads/2017/12/Internet-Citizens-ISD-Impact-Report-Dec-2017.pdf>
- Reynolds, L., et Parker, L.** (2018). *Digital Resilience Stronger Citizens Online* (p. 42). Institute for Strategic Dialogue (ISD). https://www.isdglobal.org/wp-content/uploads/2018/10/Digital-Resilience-Project-Report_-_FINAL_web.pdf
- Ris, L., et Ernstorfer, A.** (2017). *Borrowing a wheel : Applying existing design, monitoring and evaluation strategies to emerging programming approaches to prevent and counter violent extremism*. Peacebuilding evaluation consortium.
- Rodon, C.** (2018). *Rapport d'évaluation d'un dispositif / d'un programme en milieu ouvert de déradicalisation et désengagement de l'idéologie violente pour la prévention et la réduction du risque terroriste* (p. 120). ITG-Consultant. https://www.researchgate.net/publication/337495623_Rapport_d%27evaluation_d%27un_dispositif_d%27un_programme_en_milieu_ouvert_de_deradicalisation_et_desengagement_de_l%27ideologie_violente_pour_la_prevention_et_la_reduction_du_risque_terroriste_Paris_France
- Romaniuk, P.** (2015). *Does CVE work ? : Lessons learned from the global effort to counter violent extremism*. Global Center on Cooperative Security.
- Rooke, A., et Slater, I.** (2010). *Prevent in Southwark – 2009-2010 Evaluation Report*. The Centre for Urban and Community Research.
- Rothstein, H., Sutton, A. J., et Borenstein, M. (Éds.).** (2005). *Publication bias in meta-analysis : Prevention, assessment and adjustments*. Wiley.
- Rustan, E., Hanifah, N., & Kanro, B.** (2018). De-radicalization in the Implementation of Islamic Education Curriculum in SMA Masamba South Sulawesi. *Dinamika Ilmu*, 18(2), 271-283. <https://doi.org/10.21093/di.v18i2.1338>

- Sabir, R.** (2014). *Understanding Counter-Terrorism Policy and Practice in the UK since 9/11* [University of Bath]. https://purehost.bath.ac.uk/ws/portalfiles/portal/187958374/Rizwaan_Sabir_2014_Understanding_Counter_Terrorism_Policy_and_Practice_in_the_UK_since_9_11..pdf
- Salkind, N.** (2010). Quasi-Experimental Design. In *Encyclopedia of Research Design* (p. 1172-1176). SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Saltman, E. M., Dow, M., et Bjornsgaard, K.** (2016). *Youth Innovation Labs A Model for Preventing and Countering Violent Extremism* (p. 40). Institute for Strategic Dialogue (ISD). <https://www.isdglobal.org/wp-content/uploads/2016/07/YouthCAN-Labs.pdf>
- Salyk-Virk, M. J.** (2018). Building Community Resilience ? Community Perspectives of the Countering Violent Extremism Pilot Program in Minneapolis/St. Paul. *Studies in Conflict & Terrorism*, 43(11), 1011-1042. <https://doi.org/10.1080/1057610X.2018.1514054>
- Sarota, A.** (2017). *Baseline Evaluation of : Katika Usalama Tunategemeana and Pamoja ! Strengthening Community Resilience in Tanzania* (p. 107). Search for Common Ground. https://www.sfcg.org/wp-content/uploads/2017/07/Baseline-Report.Final_Public.pdf
- Savage, S., Khan, A., et Liht, J.** (2014). Preventing violent extremism in Kenya through value complexity : Assessment of Being Kenyan Being Muslim. *Journal of Strategic Security*, 7(3), 1-26. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.5038/1944-0472.7.3.1>
- Savoia, E., Su, M., Harriman, N., et Testa, M. A.** (2019). Evaluation of a School Campaign to Reduce Hatred. *Journal for Deradicalization*, 21, 43-83.
- Savoia, E., Testa, M. A., Stern, J., Lin, L., Konate, S., et Klein, N.** (2016). *Evaluation of the Greater Boston Countering Violent Extremism (CVE) Pilot Program* (p. 57). Harvard T.H. Chan School of Public Health. https://www.dhs.gov/sites/default/files/publications/OPSR_TP_CVE-Formative-Evaluation-Greater-Boston-CVE-Pilot-Program-Report_161121-508.pdf
- Schanzer, D., et Eyerman, J.** (2019). *Engaging With Communities to Prevent Violent Extremism A Review of the Obama Administration's CVE Initiative* (p. 85). Duke University and RTI International. https://www.researchgate.net/publication/336914453_Engaging_with_Communities_to_Prevent_Violent_Extremism_A_Review_of_the_Obama_Administration's_CVE_Initiative
- Schanzer, D., Kurzman, C., Toliver, J., et Miller, E.** (2016). *The Challenge and Promise of Using Community Policing Strategies to Prevent Violent Extremism : A Call for Community Partnerships with Law Enforcement to Enhance Public Safety, Final Report* (N° 249674 ; p. 87). Triangle Center on Terrorism and Homeland Security Sanford School of Public Policy, Duke University. <https://www.ncjrs.gov/pdffiles1/nij/grants/249674.pdf>
- Schmid, A. P.** (2013). Radicalisation, de-radicalisation, counter-radicalisation : A conceptual discussion and literature review. *ICCT Research Paper*, 97. <http://www.academia.edu/download/31064974/ICCT-Schmid-Radicalisation-De-Radicalisation-Counter-Radicalisation-March-2013.pdf>
- Schorn, F., Moubayed, L., et Auten, S.** (2010). *Review of the Office of Middle East Programs Youth Initiatives* (RAN-1-00-09-00016-00 ; p. 176). USAID and Aguirre Division of JBS International, Inc.
- Schulze, K. E.** (2008). Indonesia's Approach to Jihadist Deradicalization. *CTC Sentinel*, 1(8), 2.
- Schumicky-Logan, L.** (2017). Addressing Violent Extremism with a Different Approach : The Empirical Case of At-Risk and Vulnerable Youth in Somalia. *Journal of Peacebuilding & Development*, 12(2), 66-79. <https://doi.org/10.1080/15423166.2017.1336467>
- Schuurman, B.** (2018). Research on Terrorism, 2007-2016 : A Review of Data, Methods, and Authorship. *Terrorism and Political Violence*, 32(5), 1011-1026. <https://doi.org/10.1080/09546553.2018.1439023>
- Schuurman, B., et Bakker, E.** (2016). Reintegrating jihadist extremists : Evaluating a Dutch initiative, 2013-2014. *Behavioral Sciences of Terrorism and Political Aggression*, 8(1), 66-85. <https://doi.org/10.1080/19434472.2015.1100648>
- Search for Common Ground.** (2011). *Program Evaluation Report "Countering and Preventing Radicalization in Indonesian Prisons"* (p. 14). Search for Common Ground. https://www.dmeformpeace.org/sites/default/files/INA_EV_Feb11_Countering%20and%20Preventing%20Radicalization%20in%20Indonesian%20Prisons.pdf

- SecDev.Foundation.** (2016). *Extreme Dialogue : Social media Target Audience Analysis and Impact Assessments in support of countering violent extremism An abridged summary report of findings and lessons learned* (p. 41). SecDev.Foundation and The Institute for Strategic Dialogue. [https://preventviolentextremism.info/sites/default/files/Kanishka-Secdev %20Extreme %20Dialogue- %20Social %20media %20Target %20Audience %20Analysis %20and %20Impact %20Assessments %20in %20support %20of %20countering %20violent %20extremism.pdf](https://preventviolentextremism.info/sites/default/files/Kanishka-Secdev%20Extreme%20Dialogue-%20Social%20media%20Target%20Audience%20Analysis%20and%20Impact%20Assessments%20in%20support%20of%20countering%20violent%20extremism.pdf)
- Sheikh, S., Sarwar, S., et King, E.** (2012). *Evaluation of the Muslim Council of Wales' Prevent work Final report* (N° 23/2012; p. 90). Welsh Government Social Research. <https://gov.wales/sites/default/files/statistics-and-research/2019-08/120719muslimcouncilen.pdf>
- Sian, K. P.** (2015). Spies, surveillance and stakeouts : Monitoring Muslim moves in British state schools. *Race Ethnicity and Education*, 18(2), 183-201. International Bibliography of the Social Sciences (IBSS). <https://doi.org/10.1080/13613324.2013.830099>
- Silke, A.** (2001). The devil you know : Continuing problems with research on terrorism. *Terrorism and political violence*, 13(4), 1-14.
- Silke, A.** (2006). The impact of 9/11 on research on terrorism. In *Mapping terrorism research* (p. 90-107). Routledge.
- Silverman, T., Stewart, C. J., Amanullah, Z., et Birdwell, J.** (2016). *The Impact of Counter-Narratives Insights from a year-long cross-platform pilot study of counter-narrative curation, targeting, evaluation and impact* (p. 54). Institute for Strategic Dialogue (ISD) and Against Violent Extremism (AVE). https://www.isdglobal.org/wp-content/uploads/2016/08/Impact-of-Counter-Narratives_ONLINE_1.pdf
- Sjøen, M. M., et Mattsson, C.** (2019). Preventing radicalisation in Norwegian schools : How teachers respond to counter-radicalisation efforts. *Critical Studies on Terrorism*, 13(2), 218-236. <https://doi.org/10.1080/17539153.2019.1693326>
- Spalek, B., et Davies, L.** (2012). Mentoring in relation to violent extremism : A study of role, purpose, and outcomes. *Studies in Conflict and Terrorism*, 35(5), 354-368. International Bibliography of the Social Sciences (IBSS).
- Speckhard, A., Shajkovci, A., et Ahmed, M.** (2019). Intervening in and Preventing Somali-American Radicalization with Counter Narratives : Testing the Breaking the ISIS Brand Counter Narrative Videos in American Somali Focus Group Settings. *Journal of Strategic Security*, 11(4), 32-71. <https://doi.org/10.5038/1944-0472.11.4.1695>
- Speckhard, A., Shajkovci, A., Wooster, C., et Izadi, N.** (2018). Mounting a Facebook Brand Awareness and Safety Ad Campaign to Break the ISIS Brand in Iraq. *Perspectives on Terrorism*, 12(3), 50-66.
- Sullivan, H., et Stewart, M.** (2006). Who Owns the Theory of Change ? *Evaluation*, 12(2), 179-199. <https://doi.org/10.1177/1356389006066971>
- Supratno, H., Subandiyah, H., et Raharjo, R. P.** (2018). *Character Education in Islamic Boarding School as a Medium to Prevent Student Radicalism*. 222, 405-410. <https://doi.org/10.2991/soshec-18.2018.86>
- Swedberg, J.** (2011). *Mid-Term Evaluation of USAID's Counter-Extremism Programming in Africa* (p. 110). USAID. https://pdf.usaid.gov/pdf_docs/pdacr583.pdf
- Swedberg, J., et Reisman, L.** (2013). *Mid-Term Evaluation of Three Countering Violent Extremism Projects* (p. 136). USAID. https://pdf.usaid.gov/pdf_docs/pdacx479.pdf
- Taylor, E. (Lily), Taylor, P. C., Karnovsky, S., Aly, A., et Taylor, N.** (2016). "Beyond Bali" : A transformative education approach for developing community resilience to violent extremism. *Asia Pacific Journal of Education*, 37(2), 193-204. <https://doi.org/10.1080/02188791.2016.1240661>
- Taylor, L., et Soni, A.** (2017). Preventing radicalisation : A systematic review of literature considering the lived experiences of the UK's Prevent strategy in educational settings. *Pastoral Care in Education*, 35(4), 241-252. <https://doi.org/10.1080/02643944.2017.1358296>
- Tesfaye, B., McDougal, T., Maclin, B., et Blum, A.** (2018). *"If Youth Are Given the Chance" Effects of Education and Civic Engagement on Somali Youth Support of Political Violence* (p. 42). Mercy Corps. [https://www.mercycorps.org/sites/default/files/2019-11/If %20Youth %20Are %20Given %20the %20Chance_LR_FINAL.pdf](https://www.mercycorps.org/sites/default/files/2019-11/If%20Youth%20Are%20Given%20the%20Chance_LR_FINAL.pdf)
- Tesfaye, B., et Mohamud, A.** (2016). *Critical Choices Assessing the Effects of Education and Civic Engagement on Somali Youths' Propensity Towards Violence* (p. 34). Mercy Corps. https://www.mercycorps.org/sites/default/files/2020-01/CRITICAL_CHOICES_REPORT_FINAL_DIGITAL.pdf

- Thomas, P., Miah, S., et Purcell, M.** (2017). *The Kirklee Prevent Young Peoples' Engagement Team : Insights and lessons from its first year* (p. 21). Kirklees Council and the University of Huddersfield, Huddersfield Centre for Research in Education and Society (HudCRES). <http://eprints.hud.ac.uk/id/eprint/32393/1/Kirklees%20Prevent%20Engagement%20Team%20Report%20June%202017.pdf>
- Tines, J., Haq Siddiqui, Noman ul, Akhtar, N., Sadiq, M., Tanveer, T., et Iqbal Zaidi, S. Z.** (2017). *Karachi Youth Workforce Development Project : Midterm Evaluation Report (AMANTECH)* (AID-391-C-15-00004 ; p. 184). USAID.
- Tropp, L. R., Bilali, R., et Flickinger, S.** (2019). *Healing Our Communities : Promoting Social Cohesion in Rwanda* (p. 111). USAID, Karuna Center for Peacebuilding, AEGIS Preventing Crimes Against Humanity, HROC, Institute of Research and Dialogue for Peace (IRDP). <https://www.karunacenter.org/wp-content/uploads/2018/03/Healing-Our-Communities-Final-Report.pdf>
- Tsuroyya, T.** (2017). Media to Counter Radicalization : A Case Study at Islamic (Boarding) Schools. *Advanced Science Letters*, 23(12), 11649-11653. <https://doi.org/10.1166/asl.2017.10486>
- Uhlmann, M.** (2017). *Evaluation of the Advice Centre on Radicalisation Final Report* (p. 108). Federal Office for Migration and Refugees and Research Centre Migration, Integration and Asylum. https://www.beratungsstelle-radikalisierung.de/SharedDocs/Anlagen/EN/broschuere-fb31-evaluationsbericht-pdf.pdf?__blob=publicationFile&v=3
- UNEG.** (2016). *Normes et règles d'évaluation*. Groupe des Organisation des Nations unies pour l'évaluation.
- United Kingdom House of Commons, et Communities and Local Government Committee.** (2010). *Preventing Violent Extremism, Sixth Report of Session 2009-10* (Report, Together with Formal Minutes, Oral and Written Evidence HC 65 ; p. 310). House of Commons, Communities and Local Government Committee. <https://publications.parliament.uk/pa/cm200910/cmselect/cmcomloc/65/65.pdf>
- United States Government Accountability Office (GAO).** (2017). *Countering Violent Extremism, Actions Needed to Define Strategy and Assess Progress of Federal Efforts* (GAO-17-300 ; p. 62). United States Government Accountability Office (GAO). <https://www.gao.gov/assets/690/683984.pdf>
- University of Amsterdam.** (2013). *Empirical Study (revised)* (N° 241744 ; p. 127). Scientific Approach to Formulate Indicators& Responses to Radicalisation (SAFIRE) and University of Amsterdam.
- Upton, M., et Grossman, M.** (2019). The Dury's Out : Participatory drama and applied theatre processes as ways of addressing radicalized thinking – a pilot study. *Applied Theatre Research*, 7(1), 51-66. https://doi.org/10.1386/atr_00005_1
- van der Heide, L., et Schuurman, B.** (2018). Reintegrating Terrorists in the Netherlands : Evaluating the Dutch approach. *Journal for Deradicalization*, 17, 196-239.
- van Hemert, D., van der Berg, H., van Vliet, T., Roelofs, M., Huis in 't Veld, M., Marret, J.-L., Gallucci, M., et Feddes, A. R.** (2014). *Synthesis report on the state-of-the-art in evaluating the effectiveness of counter-violent extremism interventions* (D2.2). Impact Europe.
- Veldhuis, T. M.** (2015). *Captivated by fear An evaluation of terrorism detention policy* [University of Groningen].
- Veldhuis, T. M., Gordijn, E. H., Lindenberg, S. M., et Veenstra, R.** (2010). *Terrorists in Prison Evaluation of the Dutch terrorism wing* (p. 8). University of Groningen Faculty of Behavioural and Social Sciences.
- Vergani, M., Iqbal, M., Ilbahar, E., et Barton, G.** (2020). The Three Ps of Radicalization : Push, Pull and Personal. A Systematic Scoping Review of the Scientific Evidence about Radicalization Into Violent Extremism. *Studies in Conflict & Terrorism*, 43(10), 854-854. <https://doi.org/10.1080/1057610X.2018.1505686>
- Vermeulen, F.** (2014). Suspect Communities-Targeting Violent Extremism at the Local Level : Policies of Engagement in Amsterdam, Berlin, and London. *Terrorism and Political Violence*, 26(2), 286-306. Sociological Abstracts. <https://doi.org/10.1080/09546553.2012.705254>
- Vittum, K., Ombok, O., Odary, K., Mmoji, G., et Management Systems International.** (2016). *Evaluation Kenya Tuna Uwezo : Final Performance Evaluation USAID/Kenya and East Africa Office of Democracy, Governance and Conflict* (AID-623-TO-16-00004 ; p. 102). USAID and Management Systems International.
- Walsh, M., & Gansewig, A.** (2019). A former right-wing extremist in school-based prevention work : Research findings from Germany. *Journal for Deradicalization*, 21, 1-42.
- Warrington, A.** (2018). 'Sometimes you just have to try something'—A critical analysis of Danish state-led initiatives countering online radicalisation. *Journal for Deradicalization*, 14, 111-152.

Waterhouse Consulting Group. (2008). *Preventing Violent Extremism, An Independent Evaluation of the Birmingham Pathfinder* (p. 53). Waterhouse Consulting Group. <https://wallscometumblingdown.files.wordpress.com/2008/11/birmingham-pve-final-report-14-11-08.pdf>

Webb, E. (2017). *For Our Children : An Examination of Prevent in the Curriculum* (p. 32). Centre for the Response to Radicalisation and Terrorism (CRT) and The Henry Jackson Society. <https://henryjacksonsociety.org/wp-content/uploads/2018/12/For-Our-Children-An-Examination-of-Prevent-in-the-Curriculum-.pdf>

Webber, D., Chernikova, M., Kruglanski, A. W., Gelfand, M. J., Hettiarachchi, M., Gunaratna, R., Lafreniere, M.-A., et Belanger, J. J. (2018). Deradicalizing Detained Terrorists. *Political Psychology*, 39(3), 539-556. <https://doi.org/10.1111/pops.12428>

Weeks, D. (2017). Doing Derad : An Analysis of the U.K. System. *Studies in Conflict & Terrorism*, 41(7), 523-540. <https://doi.org/10.1080/1057610X.2017.1311107>

Weinberg, L., Pedahzur, A., et Hirsch-Hoefler, S. (2004). The Challenges of Conceptualizing Terrorism. *Terrorism and Political Violence*, 16(4), 777-794. <https://doi.org/10.1080/095465590899768>

Weine, S., Eisenman, D., Glik, D., Kinsler, J., et Polutnik, C. (2016). *Leveraging a Targeted Violence Prevention Program to Prevent Violent Extremism*: (p. 34). University of Illinois at Chicago (UIC) and University of California, Los Angeles (UCLA). https://www.dhs.gov/sites/default/files/publications/862_OPSR_TP_LA-Formative-Evaluation_180817-508.pdf

Widmer, T., Blaser, C., et Falk, C. (2007). Evaluating Measures Taken Against Right-Wing Extremism. *Evaluation*, 13(2), 221-239. ProQuest Sociology Collection. <https://doi.org/10.1177/1356389007075225>

Wilchen Christensen, T. (2015). *A Question of Participation : Disengagement from the Extreme Right. A case study from Sweden* [Kopicalentralen, Roskilde Universitet]. https://rucforsk.ruc.dk/ws/portalfiles/portal/56384428/twc_fin_ny.pdf

Williams, M. J., Horgan, J. G., et Evans, W. P. (2016). *Evaluation of a Multi-Faceted, U.S. Community-Based, Muslim-Led CVE Program* (N° 249936 ; p. 167). Georgia State University and University of Nevada, Reno. <https://www.ncjrs.gov/pdffiles1/nij/grants/249936.pdf>

Williams, M. J., et Kleinman, S. M. (2014). A utilization-focused guide for conducting terrorism risk reduction program evaluations. *Behavioral Sciences of Terrorism and Political Aggression*, 6(2), 102. ProQuest Sociology Collection.

Wilner, A., et Rigato, B. (2017). The 60 Days of PVE Campaign : Lessons on Organizing an Online, Peer-to-Peer, Counter-radicalization Program. *Journal for Deradicalization*, 12, 227-268.

Wilson, N. L., et Krentel, J. (2018). *Lessons from Strengthening Capacity in Countering Violent Extremism* (p. 20). United States Institute of Peace. <https://www.usip.org/publications/2018/05/lessons-strengthening-capacity-countering-violent-extremism>

Winston, J., et Strand, S. (2013). Tapestry and the aesthetics of theatre in education as dialogic encounter and civil exchange. *Research in Drama Education: The Journal of Applied Theatre and Performance*, 18(1), 62-78. <https://doi.org/10.1080/13569783.2012.756178>

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., et Hasisi, B. (2019). A Field-Wide Systematic Review and Meta-analysis of Putative Risk and Protective Factors for Radicalization Outcomes. *Journal of Quantitative Criminology*. <https://doi.org/10.1007/s10940-019-09439-4>

Young, H., Rooze, M., Russell, J., Ebner, J., et Schulten, N. (2016). *Evidence-based Policy Advice Final Report* (p. 54). Terrorism and Radicalisation (TerRa).

Younis, T., et Jadhav, S. (2019). Keeping Our Mouths Shut : The Fear and Racialized Self-Censorship of British Healthcare Professionals in PREVENT Training. *Cult Med Psychiatry*, 43, 404-424. <https://doi.org/10.1007/s11013-019-09629-6>

Zeuthen, M. (2021). *Reintegration : Disengaging Violent Extremists - A Systematic Literature Review of Effectiveness of Counter-Terrorism and Preventing and Countering Violent Extremism Activities*. Rusi. https://english.iob-evaluatie.nl/binaries/iob-evaluatie-eng/documenten/sub-studies/2021/02/01/literature-studies-%E2%80%93-counterterrorism-and-preventing-and-countering-violent-extremism/Rusi_Reintegration_disengaging_violent_extremists_202102.pdf

Annexe A : Liste des études retenues

Tableau 31. Études retenues

Étude (études fusionnées) ⁵¹	Pays	Prévention	Extrémisme ciblé	Action ciblée	Type d'évaluation	Design global	Type d'étude	Prépost	Participants	Groupe témoin
Abu-Nimer et Nasser, 2017	Niger	primaire	islamiste	projet	processus	qualitative	observationnelle	post évaluation	46	
Admo et al., 2018	Canada	primaire ciblée	non spécifique	projet	impact, d'extrants	mixte	observationnelle	post évaluation	102	
Aldrich, 2012	Mali, Chad, Niger	primaire	non spécifique	projet	impact	quantitative	quasi expérimentale	post évaluation	>1000	
Aldrich, 2014	Mali	primaire	islamiste	projet	impact	quantitative	quasi expérimentale	post évaluation	200	
Algristian et al., 2019	Indonésie	générale	non spécifique	projet	impact	quantitative	quasi expérimentale	prépost	16	
Ali et Saragih, 2018	Indonésie	générale	non spécifique	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	Non informé	
Al-Maqosi et al., 2019	Jordanie	primaire ciblée	islamiste	projet	impact	quantitative	expérimentale	prépost	23	25
Amanullah et Harrasy, 2017	Kenya	secondaire	islamiste	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Anindya, 2019	Indonésie	secondaire	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	21	
Audit Commission, 2008	Royaume-Uni	primaire ciblée	non spécifique	plan ou stratégie nationale	impact, audit	qualitative	observationnelle	post évaluation	Non informé	
Awan, 2012	Royaume-Uni	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	6	
Azam et Bareeha, 2017	Pakistan	tertiaire	islamiste	projet	impact, processus	qualitative	observationnelle	post évaluation	67	
Badurdeen et Goldsmith, 2018	Kenya	générale	islamiste	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	249	
Bala et Deman, 2017	Tunisie	primaire ciblée	non spécifique	projet	impact, processus, d'extrants, autre	qualitative	observationnelle	post évaluation	516	

⁵¹ Lorsqu'il y a lieu, la référence des articles fusionnés est indiquée entre parenthèses.

Barkindo et Bryans, 2016	Nigéria	tertiaire	non spécifique	composante du plan ou de la stratégie nationale	processus	qualitative	N/A	post évaluation	Non informé	
Basse, 2018	Niger	secondaire	islamiste	projet	impact, processus, d'extrants	mixte	observationnelle	post évaluation	201	
Bastug et Evlek, 2016	Turquie	secondaire, tertiaire	extrême gauche, extrême droite, islamiste, autre	composante du plan ou de la stratégie nationale	impact	quantitative	observationnelle	post évaluation	Non informé	
Bean et al., 2011	Chad, Niger	primaire	non spécifique	projet	impact, processus	qualitative	observationnelle	post évaluation	Impossible à déterminer	
Bilali, 2019	Burkina Faso	primaire	non spécifique	projet	impact	mixte	expérimentale	prépost	1 452	1452
Bilazarian, 2016	Royaume-Uni	secondaire, tertiaire	non spécifique	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	4	
Bou Zeid, 2019	Liban	primaire ciblée	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	5	
Boucek, 2008 (Boucek, 2009)	Arabie Saoudite	tertiaire	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	5	
Boulton, 2010	Philippines	générale	non spécifique	projet	autre	qualitative	observationnelle	post évaluation	4	
Bowie et Revell, 2018	Royaume-Uni	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	8	
Boyd-MacMillan, 2016	Écosse	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	mixte	quasi expérimental	prépost	21	
Boyle et al., 2016	Maroc	primaire ciblée, secondaire	non spécifique	projet	impact, autre	mixte	observationnelle	post évaluation	6	
Brett et Kahlmeyer, 2017	Kenya / Somalie	générale	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Briggs, 2010	Royaume-Uni	primaire ciblée, secondaire	non spécifique	composante du plan ou de la stratégie nationale	impact	mixte	observationnelle	post évaluation	77	
Broadbent, 2013	Australie	secondaire	non spécifique	projet	impact, processus	quantitative	observationnelle	post évaluation	16	
Brooks et Ezzani, 2017	États-Unis	primaire ciblée	islamiste	projet	processus, autre	qualitative	observationnelle	post évaluation	Impossible à déterminer	
Brottsförebyggande rådet (Sweden), 2001 ⁵²	Suède	secondaire, tertiaire	extrême droite	plan ou stratégie nationale	impact, processus, d'extrants	qualitative	observationnelle	post évaluation	Non informé	

⁵² Sommaire en anglais.

Bryan, 2017	Royaume-Uni	secondaire	extrême droite, islamiste	composante du plan ou de la stratégie nationale	processus, d'extrants	qualitative	observationnelle	post évaluation	3	
Busher et al., 2017	Royaume-Uni	primaire	non spécifique	composante du plan ou de la stratégie nationale	processus	mixte	observationnelle	post évaluation	303	
Chatellier, 2012	Pakistan	primaire ciblée	non spécifique	projet	impact, processus	mixte	observationnelle	post évaluation	Non informé	
Cherney et Belton, 2019 (Cherney, 2020)	Australie	secondaire, tertiaire	non spécifique	projet	impact, processus	mixte	observationnelle	post évaluation	14-22	
Christiaens et al., 2018	Pays-Bas	générale	non spécifique	projet	impact, processus	mixte	quasi expérimentale	Post évaluation avec suivi	101	
Christmann et al., 2012	Royaume-Uni	générale	islamiste	composante du plan ou de la stratégie nationale	impact, processus, d'extrants, suivi	mixte	observationnelle	post évaluation	33-77	
Cifuentes et al., 2013	Royaume-Uni	primaire ciblée, secondaire	extrême droite	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Cipaku, 2013	Indonésie	tertiaire	islamiste	projet	impact, d'extrants	qualitative	observationnelle	post évaluation	45	
Clemens-Hope, 2015	Niger, Chad et Burkina Faso	primaire	non spécifique	projet	d'extrants	autre	observationnelle	post évaluation	Non informé	
Cockayne et al., 2015	Somali et Kenya	tertiaire	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	67	
Colibaba et al., 2017	Roumanie	générale	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	40	
Court, 2006	Israël	primaire ciblée	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Cragin et Chalk, 2003a	Irlande du Nord	primaire	non spécifique	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	Non informé	
Cragin et Chalk, 2003b	Philippines	primaire	autre, non spécifique	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	Non informé	
Cragin et Chalk, 2003c	Israël	primaire	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Davey et al., 2018	n/a	tertiaire	extrême droite, islamiste	projet	impact	mixte	observationnelle	post évaluation	> 800	
Demant et al., 2009	Pays-Bas	secondaire, tertiaire	extrême droite	projet	impact	qualitative	observationnelle	post évaluation	22	
Dhungana et al., 2016	Kyrgystan et Asie centrale	primaire	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	48	
Dietrich, 2018	Nigéria	primaire, primaire ciblée	islamiste	projet	impact, processus	mixte	observationnelle	post évaluation	1282	

Dunn et al., 2015	Australie	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact, processus, autre	mixte	observationnelle	post évaluation	33	
Dwyer et Maruna, 2011 (Dwyer, 2010)	Irlande du Nord	tertiaire	autre	projet	processus	mixte	observationnelle	post évaluation	104	
Education Development Center (EDC) et USAID, 2019a	Philippines	primaire ciblée	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	1657	41
Education Development Center (EDC) et USAID, 2019b	Philippines	primaire ciblée	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	789	
Eriksson, 2008	Irlande du Nord	primaire	autre	projet	processus, autre	qualitative	observationnelle	post évaluation	Non informé	
Feddes et al., 2019a	Pays-Bas	primaire ciblée	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	228	
Feddes et al., 2019b	Pays-Bas	primaire ciblée	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	225	
Finkel et al., 2015a	Chad	primaire, secondaire	non spécifique	projet	impact, d'extrants, autre	mixte	quasi expérimentale	prépost (avec évaluation intermédiaire)	450	15
Finkel et al., 2015b	Niger	primaire, secondaire	non spécifique	projet	impact, d'extrants, autre	mixte	quasi expérimentale	prépost (avec évaluation intermédiaire)	450	15
Finkel et al., 2018a (Finkel et al., 2017)	Niger	primaire, secondaire	non spécifique	projet	impact, d'extrants, autre	mixte	quasi expérimentale	prépost (avec évaluation intermédiaire)	18185	
Finkel et al., 2018b (Finkel et al., 2017)	Chad	primaire, primaire ciblée	non spécifique	projet	impact	quantitative	expérimentale	Post évaluation avec suivi	18185	
Finkel et al., 2018c (Finkel et al., 2017)	Burkina Faso	primaire, primaire ciblée	non spécifique	projet	impact	quantitative	expérimentale	Post évaluation avec suivi	18185	
Finn et al., 2016	Kenya	générale	non spécifique	plan ou stratégie nationale	impact	qualitative	observationnelle	post évaluation	Non informé	
Franssen et al., 2019	Belgique	primaire, secondaire, tertiaire	non spécifique	projet	impact, processus, d'extrants	mixte	observationnelle	post évaluation	945	
Frenett et Dow, 2015	n/a	secondaire, tertiaire	extrême droite, islamiste	projet	impact	mixte	observationnelle	post évaluation	154	
Garaigordobil, 2012	Espagne	générale	autre	projet	impact	quantitative	quasi expérimental	prépost	191	85
Gatewood et Boyer, 2019	France	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimentale	prépost	22	
Glazzard et Reed, 2018	n/a	générale	non spécifique	composante du plan ou de la stratégie nationale	processus, d'extrants	qualitative	observationnelle	post évaluation	21	
Goaziou, 2018	France	secondaire, tertiaire	islamiste, non spécifique		processus	qualitative	N/A	post évaluation	Impossible à déterminer	1
Government of the United Kingdom, 2011	Royaume-Uni	primaire ciblée, secondaire	islamiste	plan ou stratégie nationale	impact, processus	mixte	observationnelle	post évaluation	1113	
Greiner, 2010	Niger et Chad	primaire	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	182	

Harahap et al., 2019	Indonésie	secondaire	islamiste	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Harris-Hogan et al., 2019	Australie	générale	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	117	
Heath-Kelly et Strausz, 2018	Royaume-Uni	primaire	non spécifique	composante du plan ou de la stratégie nationale	impact, processus, autre	mixte	observationnelle	post évaluation	335	
Helmus et Klein, 2018	n/a	secondaire	extrême droite, islamiste	projet	impact	quantitative	observationnelle	post évaluation	Non informé	
Hiariej et al., 2017	Indonésie	secondaire, tertiaire	islamiste	composante du plan ou de la stratégie nationale	impact, processus	mixte	observationnelle	post évaluation	1170	
Hirschi et Widmer, 2012a	Suisse	primaire	extrême droite	projet	impact	quantitative	quasi expérimental	prépost	Non informé	
Hirschi et Widmer, 2012d	Suisse	primaire	extrême droite	projet	impact	quantitative	quasi expérimental	prépost	747	12
Hirschi et Widmer, 2012e	Suisse	primaire ciblée, secondaire	extrême droite	projet	impact	quantitative	observationnelle	post évaluation	115	
Hirschi et Widmer, 2012g	Suisse	Primaire, primaire ciblée, secondaire	extrême droite	projet	impact	quantitative	quasi expérimental	prépost	Non informé	
Iacopini et al., 2011	Royaume-Uni	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	37	
Ipp et al., 2014	Tunisie	secondaire	islamiste	composante du plan ou de la stratégie nationale	impact, autre	qualitative	observationnelle	post évaluation	14	
Istiqomah, 2011	Indonésie	tertiaire	islamiste	projet	impact	qualitative	observationnelle	post évaluation	4	
i-works research ltd., 2013	Wales	secondaire	extrême droite	composante du plan ou de la stratégie nationale	impact	mixte	quasi expérimental	prépost	Non informé	
Jackson et al., 2019	États-Unis	générale	non spécifique	composante du plan ou de la stratégie nationale	autre	qualitative	observationnelle	post évaluation	50	
Jailobaeva et Asilbekova, 2017	Kirghizistan	tertiaire	islamiste	projet	impact, processus, d'extrants	mixte	observationnelle	post évaluation	41	
Jerome et Elwick, 2016 (Jerome et Elwick, 2019 ; Elwick et Jerome, 2019)	Royaume-Uni	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimental	prépost	232	
Johns et al., 2014	Australie	primaire ciblée	islamiste	projet	impact	mixte	observationnelle	post évaluation	39	
Johnston et al., 2008a	Pakistan	primaire	islamiste	projet	impact	mixte	observationnelle	post évaluation	Non informé	
Johnston et al., 2008b	Pakistan	primaire	islamiste	projet	impact	mixte	observationnelle	post évaluation	Non informé	
Joyce, 2018	Royaume-Uni	secondaire	non spécifique	composante du plan ou de la stratégie nationale	processus	mixte	observationnelle	post évaluation	38	

Khalil et Ipp, 2016	Mali	primaire	non spécifique	projet	impact, autre	qualitative	observationnelle	post évaluation	Impossible à déterminer	
Khalil et Zeuthen, 2014	Kenya	primaire	non spécifique	projet	autre	qualitative	observationnelle	post évaluation	Non informé	
Khalil et al., 2019	Somalie	tertiaire	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	102	
Khurshid et al., 2018	Pakistan	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	impact	mixte	observationnelle	post évaluation	500	
Kollmorgen et Barry, 2017	Thaïlande	générale	non spécifique	projet	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Kollmorgen et al., 2019	Kenya	primaire	islamiste	composante du plan ou de la stratégie nationale	impact, processus, d'extrants	mixte	observationnelle	post évaluation	528	
Kundnani, 2009	Royaume-Uni	primaire ciblée, secondaire, tertiaire	islamiste	plan ou stratégie nationale	processus, d'extrants	qualitative	observationnelle	post évaluation	56	
Kurtz, 2015 (Kurtz et al., 2016)	Afghanistan	secondaire	non spécifique	projet	impact, processus	mixte	quasi expérimentale	post évaluation	1129	
Kyriacou et al., 2017	Royaume-Uni	primaire, primaire ciblée	islamiste, non spécifique	plan ou stratégie nationale	impact, autre	mixte	observationnelle	post évaluation	9	
L. Parker et al., 2018	Italie, Roumanie et Suède	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimental	prépost	192	135
Lakhani, 2012	Royaume-Uni	primaire ciblée	islamiste, non spécifique	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	56	
Lamhaidi, 2017	Maroc	secondaire	non spécifique	projet	impact, processus, d'extrants	mixte	observationnelle	post évaluation	Impossible à déterminer	
Letsch, 2018	Tunisie	générale	islamiste, non spécifique	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	25	
Levy et al., 2019	Kirghizistan	secondaire	islamiste, non spécifique	projet	impact, processus	mixte	observationnelle	post évaluation	1644	
Liht et Savage, 2013	Royaume-Uni	primaire ciblée	islamiste	projet	impact	quantitative	quasi expérimental	prépost	81	
Lindekilde, 2012	Danemark	primaire ciblée, secondaire, tertiaire	non spécifique	plan ou stratégie nationale	processus, autre	qualitative	observationnelle	post évaluation	17	
Lindekilde, 2014	Danemark	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	Non informé	
Lobnikar et al., 2019	Croatie	générale	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	108	
Madriaza et al., 2018	France	primaire, secondaire, tertiaire	non spécifique	projet	impact, processus	mixte	quasi expérimentale	prépost	15-81	
Manby, 2009a	Royaume-Uni	primaire ciblée	non spécifique	projet	impact	mixte	quasi expérimental	prépost	5	

Manby, 2009b	Royaume-Uni	secondaire	non spécifique	projet	impact	mixte	quasi expérimental	prépost	9	
Manby, 2009c	Royaume-Uni	secondaire	non spécifique	projet	impact	mixte	quasi expérimental	prépost	7	
Manby, 2009d	Royaume-Uni	secondaire	extrême droite, islamiste, non spécifique	projet	impact	mixte	quasi expérimental	prépost	6	
Manby, 2010a	Royaume-Uni	secondaire, tertiaire	non spécifique	projet	impact	mixte	quasi expérimentale	prépost	5	
Manby, 2010b	Royaume-Uni	tertiaire	non spécifique	projet	impact	mixte	quasi expérimental	prépost	9	
Mansour, 2017	Maroc international	primaire	non spécifique	composante du plan ou de la stratégie nationale	impact, d'extrants	mixte	observationnelle	post évaluation	57	
Mastroe, 2016	Royaume-Uni	primaire	non spécifique	plan ou stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	20	
McDonald et Mir, 2011	Royaume-Uni	primaire ciblée	islamiste	projet	processus, autre	qualitative	observationnelle	post évaluation	48	
McDowell-Smith et al., 2017	États-Unis	primaire ciblée	islamiste	projet	impact	quantitative	observationnelle	post évaluation	75	
McGlynn et McDaid, 2016	Royaume-Uni	primaire	non spécifique	composante du plan ou de la stratégie nationale	impact, autre	qualitative	observationnelle	post évaluation	11	
McRae, 2010 (McRae, 2009a ; McRae, 2009b)	Indonésie	tertiaire	islamiste	projet	impact, processus	mixte	observationnelle	post évaluation	Non informé	
Meringolo et al., 2019a	Italie	primaire ciblée	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	18	
Meringolo et al., 2019b	Italie	primaire ciblée	non spécifique	projet	impact	mixte	observationnelle	post évaluation	19	
Mitts, 2017	États-Unis	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	quantitative	quasi expérimentale	prépost avec séries temporelles	Non informé	
Moffett et Sgro, 2016	n/a	primaire	non spécifique	projet	d'extrants	quantitative	observationnelle	post évaluation	Non informé	
Monzani et al., 2018	Kenya	générale	islamiste, non spécifique	projet	impact, autre	mixte	quasi expérimentale	prépost	194	145
Muncy et al., 2015	Nigéria	primaire	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	191	
Murtaza et al., 2018	Pakistan	secondaire	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	102	
Nicolls et Hassan, 2014	Somalie	primaire ciblée	non spécifique	projet	processus, d'extrants	qualitative	observationnelle	post évaluation	357	
Octavia et Wahyuni, 2014	Indonésie	secondaire, tertiaire	extrême gauche, extrême droite, islamiste	projet	processus	qualitative	observationnelle	post évaluation	Impossible à déterminer	
Onyima, 2017	Nigéria	primaire ciblée	islamiste	plan ou stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	68	
Orban, 2019	Norvège	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	17	

O'Toole et al., 2012	Royaume-Uni	primaire ciblée	islamiste	projet	impact, processus, d'extrants	mixte	quasi expérimental	prépost	Non informé	
O'Toole et al., 2013 (O'Toole et al., 2016)	Royaume-Uni	tertiaire	islamiste	projet	impact, processus	qualitative	observationnelle	post évaluation	112	
Parker et Lindekilde, 2020 ⁵³	Danemark	primaire ciblée	non spécifique	projet	impact	quantitative	expérimentale	prépost	955	976
Peracha et al., 2016	Pakistan	tertiaire	islamiste	projet	impact, processus	qualitative	observationnelle	post évaluation	4	
Peterson, 2012	Suède	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	13	
Piasecka, 2019	Royaume-Uni	primaire ciblée	extrême droite, islamiste	projet	impact, autre	qualitative	observationnelle	post évaluation	Non informé	
Pickering et al., 2008	Australie	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	processus	mixte	observationnelle	post évaluation	601	
Pipe et al., 2016	Somalie	primaire	non spécifique	projet	impact, autre	mixte	observationnelle	post évaluation	2789	
Powers, 2015	Royaume-Uni	primaire, primaire ciblée	non spécifique	plan ou stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	95	
Ranstorp, 2010	Indonésie	générale	islamiste	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	Non informé	
Reeves et Crowther, 2019	Royaume-Uni	primaire ciblée	non spécifique	projet	impact	quantitative	observationnelle	post évaluation	146	
Reynolds, 2017	Royaume-Uni	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimental	prépost	441	
Reynolds et Parker, 2018	n/a	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimental	prépost	54	51
Rodon, 2018	France	secondaire, tertiaire	non spécifique	projet	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Rooke et Slater, 2010	Royaume-Uni	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact, processus, d'extrants	qualitative	observationnelle	post évaluation	Non informé	
Rustan et al., 2018	Indonésie	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Sabir, 2014	Royaume-Uni	primaire ciblée, secondaire, tertiaire	islamiste	plan ou stratégie nationale	autre	qualitative	observationnelle	post évaluation	20	
Saltman et al., 2016	n/a	primaire ciblée	non spécifique	projet	impact	mixte	quasi expérimental	prépost	Impossible à déterminer	
Salyk-Virk, 2018	États-Unis	générale	non spécifique	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	26	

⁵³ Un manuscrit fut envoyé par les auteurs, en 2019.

Sarota, 2017	Tanzanie	secondaire	islamiste	projet	impact, processus	mixte	observationnelle	post évaluation	391	
Savage et al., 2014	Kenya	primaire ciblée	islamiste	projet	impact	mixte	quasi expérimental	prépost	24	
Savoia et al., 2016	États-Unis	générale	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	52	
Savoia et al., 2019	États-Unis	primaire	non spécifique	projet	impact	quantitative	quasi expérimental	prépost	767	326
Schanzer et Eyerman, 2019	États-Unis	primaire ciblée, générale	islamiste, non spécifique	plan ou stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Schanzer et al., 2016	États-Unis	primaire ciblée, secondaire	non spécifique	composante du plan ou de la stratégie nationale	processus	mixte	observationnelle	post évaluation	382 Départements de police	
Schorn et al., 2010a	Égypte	primaire ciblée	non spécifique	projet	impact, processus, autre	mixte	observationnelle	post évaluation	73	
Schorn et al., 2010b	Égypte	primaire ciblée	non spécifique	projet	impact, processus, autre	mixte	observationnelle	post évaluation	48	
Schorn et al., 2010c	Égypte	primaire ciblée	non spécifique	projet	impact, processus, autre	mixte	observationnelle	post évaluation	15	
Schulze, 2008	Indonésie	tertiaire	islamiste	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Schumicky-Logan, 2017	Somalie	secondaire, tertiaire	non spécifique	projet	impact	mixte	quasi expérimental	prépost	392	
Schuurman et Bakker, 2016	Pays-Bas	tertiaire	islamiste	projet	impact, processus	qualitative	observationnelle	post évaluation	6	
Search for Common Ground, 2011	Indonésie	tertiaire	non spécifique	projet	impact, processus	mixte	quasi expérimental	prépost	Impossible à déterminer	
SecDev.Foundation, 2016	n/a	primaire	non spécifique	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Sheikh et al., 2012	Wales	primaire ciblée, secondaire, tertiaire	islamiste	composante du plan ou de la stratégie nationale	impact, processus, d'extrants	mixte	observationnelle	post évaluation	65	
Sian, 2015	Royaume-Uni	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	Non informé	
Silverman et al., 2016a	États-Unis	primaire ciblée	islamiste	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Silverman et al., 2016b	États-Unis	primaire ciblée	islamiste	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Silverman et al., 2016c	Pakistan	tertiaire	extrême droite	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Sjøen et Mattsson, 2019	Norvège	secondaire	non spécifique	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	16	
Spalek et Davies, 2012	Royaume-Uni	secondaire	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	16	

Speckhard et al., 2018	Iraq	primaire, secondaire	islamiste	projet	impact	mixte	observationnelle	post évaluation	Impossible à déterminer	
Speckhard et al., 2019	États-Unis	secondaire	islamiste	projet	autre	mixte	observationnelle	post évaluation	Impossible à déterminer	
Supratno et al., 2018	Indonésie	primaire ciblée	islamiste	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Swedberg, 2011a	Niger	générale	non spécifique	projet	impact	mixte	quasi expérimentale	post évaluation	217	117
Swedberg, 2011b	Tchad	générale	non spécifique	projet	impact	mixte	quasi expérimentale	post évaluation	368	152
Swedberg, 2011c	Mali	générale	non spécifique	projet	impact	mixte	quasi expérimentale	post évaluation	100	1
Swedberg et Reisman, 2013	Kenya	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	mixte	quasi expérimentale	post évaluation	962	484
Taylor et al., 2016 (Aly et al., 2014)	Indonésie	primaire ciblée	non spécifique	projet	impact	mixte	observationnelle	post évaluation	21	
Tesfaye et Mohamud, 2016	Somalie	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	mixte	quasi expérimentale	post évaluation	504	298
Tesfaye et al., 2018	Somalie	primaire	islamiste	composante du plan ou de la stratégie nationale	impact	mixte	quasi expérimentale	post évaluation	937	283
Thomas et al., 2017a	Royaume-Uni	primaire ciblée	non spécifique	projet	processus	mixte	observationnelle	post évaluation	11	
Thomas et al., 2017b	Royaume-Uni	primaire ciblée	non spécifique	projet	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Tines et al., 2017	Pakistan	secondaire	non spécifique	projet	impact	mixte	quasi expérimental	prépost	801	
Tropp et al., 2019a	Rwanda	secondaire	non spécifique	projet	impact	quantitative	quasi expérimentale	post évaluation	26	27
Tropp et al., 2019b	Rwanda	secondaire	non spécifique	projet	impact, d'extrants, autre	quantitative	quasi expérimental	prépost (avec suivi)	68	
Tsuroyya, 2017	Indonésie	secondaire	islamiste	composante du plan ou de la stratégie nationale	impact, autre	qualitative	observationnelle	post évaluation	4	
Uhlmann, 2017	Allemagne	secondaire, tertiaire	non spécifique	composante du plan ou de la stratégie nationale	processus, d'extrants	qualitative	observationnelle	post évaluation	Non informé	
United Kingdom House of Commons et Communities and Local Government Committee, 2010	Royaume-Uni	primaire ciblée	islamiste	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	33	
United States Government Accountability Office (GAO), 2017	États-Unis	générale	non spécifique	plan ou stratégie nationale	d'extrants, audit	autre	observationnelle	post évaluation	6	
University of Amsterdam, 2013 (Feddes et al., 2015)	Pays-Bas	primaire ciblée	islamiste	projet	impact	mixte	quasi expérimental	prépost (avec suivi)	46	
Upton et Grossman, 2019	Australie	primaire ciblée	islamiste	projet	impact	mixte	observationnelle	post évaluation	41	

Van der Heide et Schuurman, 2018	Pays-Bas	tertiaire	non spécifique	projet	impact	qualitative	observationnelle	post évaluation	72	
Veldhuis, 2015	Pays-Bas	tertiaire	non spécifique	composante du plan ou de la stratégie nationale	processus	mixte	observationnelle	post évaluation	Non informé	
Veldhuis et al., 2010	Pays-Bas	tertiaire	non spécifique	composante du plan ou de la stratégie nationale	processus, autre	qualitative	observationnelle	post évaluation	Non informé	
Vermeulen, 2014a	Royaume-Uni	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	12	
Vermeulen, 2014b	Allemagne	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	12	
Vermeulen, 2014c	Pays-Bas	primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	12	
Vittum et al., 2016	Kenya	secondaire	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	47	
Walsh et Gansewig, 2019	Allemagne	primaire ciblée	extrême droite	projet	impact, processus	mixte	expérimentale	prépost	564	
Warrington, 2018	Danemark	primaire	non spécifique	composante du plan ou de la stratégie nationale	autre	qualitative	observationnelle	post évaluation	Non informé	
Waterhouse Consulting Group, 2008	Royaume-Uni	primaire ciblée	islamiste	composante du plan ou de la stratégie nationale	impact	qualitative	observationnelle	post évaluation	Impossible à déterminer	
Webb, 2017	Royaume-Uni	primaire	non spécifique	plan ou stratégie nationale	processus, autre	mixte	observationnelle	post évaluation	35	
Webber et al., 2018 (Kruglanski et al., 2014)	Sri Lanka	tertiaire	autre	projet	impact	quantitative	quasi expérimental	prépost (avec suivi)	669	255
Weeks, 2017	Royaume-Uni	tertiaire	non spécifique	composante du plan ou de la stratégie nationale	impact, processus	qualitative	observationnelle	post évaluation	23	
Weine et al., 2016	États-Unis	primaire	non spécifique	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Wilchen Christensen, 2015	Norvège	secondaire	extrême droite	composante du plan ou de la stratégie nationale	processus	qualitative	observationnelle	post évaluation	Non informé	
Williams et al., 2016	États-Unis	primaire ciblée, secondaire	islamiste	projet	impact, processus	mixte	quasi expérimentale	post évaluation	323	46
Wilner et Rigato, 2017	Canada	primaire	non spécifique	projet	d'extrants, autre	mixte	observationnelle	post évaluation	Non informé	
Wilson et Krentel, 2018	Émirats Unis	primaire ciblée	non spécifique	projet	impact, processus	mixte	quasi expérimentale	post évaluation avec suivi	Non informé	

Winston et Strand, 2013	Royaume-Uni	primaire ciblée	non spécifique	projet	impact, processus	mixte	observationnelle	post évaluation	260	
Young et al., 2016a	Allemagne	primaire ciblée	non spécifique	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	Non informé	
Young et al., 2016b	Allemagne	secondaire, tertiaire	non spécifique	plan ou stratégie nationale	processus	qualitative	observationnelle	post évaluation	Non informé	
Young et al., 2016c	Danemark	secondaire, tertiaire	islamiste	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Young et al., 2016d	Royaume-Uni	tertiaire	extrême droite	projet	processus	qualitative	observationnelle	post évaluation	Non informé	
Young et al., 2016e	Pays-Bas	tertiaire	islamiste	projet	impact, processus	qualitative	observationnelle	post évaluation	Non informé	
Younis et Jadhav, 2019	Royaume-Uni	primaire, primaire ciblée	non spécifique	composante du plan ou de la stratégie nationale	autre	qualitative	observationnelle	post évaluation	16	



Annexe B :

Méthode complète

La stratégie de revue systématique est basée sur les méthodes de revue du Campbell Collaboration (<https://www.campbellcollaboration.org>). Nous définissons une revue systématique comme « une revue d'une question clairement formulée qui utilise des méthodes systématiques et explicites pour identifier, sélectionner et évaluer de façon critique les recherches pertinentes, et pour recueillir et analyser les données des études qui sont incluses dans la revue » (Moher et al., 2009, p. 1). Nous avons notamment utilisé le document « *Methodological expectations of Campbell Collaboration intervention reviews: Conduct standards* » et le « PRISMA Statement » afin d'élaborer la stratégie de recherche. La méthode de la Campbell Collaboration est considérée de nos jours comme le modèle en matière de revues systématiques, particulièrement dans le domaine des sciences sociales et humaines.

B1 OBJECTIF, QUESTIONS DE RECHERCHE ET DÉFINITIONS

a) Objectif

Le but initial et global de cette revue systématique est de répertorier les évaluations de programmes de prévention de l'extrémisme violent (PEV), publiées jusqu'en décembre 2019.

En plus de cet objectif général, quatre objectifs spécifiques sont pertinents :

1. Identifier les méthodologies utilisées dans l'évaluation des programmes de prévention de l'extrémisme violent.
2. Identifier les lacunes de la littérature sur l'évaluation de ces programmes.
3. Évaluer la qualité méthodologique des études évaluatives existantes dans le domaine.
4. Formuler des recommandations pour l'évaluation de ces programmes.

b) Question de recherche

La question principale est, par conséquent, la suivante : Quelles sont les principales recommandations, en matière d'évaluation de programmes de prévention dans le domaine de l'extrémisme violent, que l'on peut tirer de la littérature ? Cette question principale implique des sous-questions associées aux concepts clés spécifiques.

Questions clés spécifiques :

- 1) Quels sont les programmes de prévention de niveau primaire qui ont été évalués ?
- 2) Quels sont les programmes de prévention de niveau secondaire qui ont été évalués ?
- 3) Quels sont les programmes de prévention de niveau tertiaire qui ont été évalués ?
- 4) Quels autres programmes, non inclus dans les trois niveaux de prévention, ont également été évalués ?
- 5) Quelles recommandations pourraient être considérées, en tenant compte à la fois du point de vue des personnes praticiennes et chercheuses impliquées dans les recherches recensées ?

Pour chaque étude, les sous-questions spécifiques suivantes ont été posées :

- a) Quelle approche théorique d'évaluation a été utilisée ?
- b) Quelle méthode d'évaluation a été utilisée ?
- c) Quels stratégies, outils et indicateurs pour mettre en œuvre une évaluation ont été utilisés ?
- d) Comment les résultats de ces programmes ont-ils été définis et mesurés ?
- e) Quelle a été la population cible ?
- f) Quelle méthode a été utilisée pour évaluer la qualité de l'évaluation ?

c) Définitions clés

Nous faisons la différence entre *radicalisation* et *radicalisation menant à la violence* (RV) en suivant les pistes données par Schmid (2013). La radicalisation est un processus dynamique qui découle de la polarisation progressive des idéaux politiques, économiques, sociaux ou religieux qui visent à rejeter ou à miner le statu quo. Elle peut conduire à des résultats positifs et négatifs pour les individus et la société, y compris des possibilités de changement social ou l'exacerbation d'un climat de confrontation entre personnes actrices ou groupes. Lorsque les moyens préconisés pour parvenir à la solution radicale impliquent de légitimer le recours à la violence ou d'envisager le passage à l'acte violent, il s'agit d'une radicalisation menant à la violence. Schmid considère que la radicalisation peut, par conséquent, être bel et bien démocratique ; en revanche, l'extrémisme implique la « tendance à ne pas respecter l'État de droit et à rejeter le pluralisme dans la société » (Schmid, 2013, p. 8). Il n'y a

pas un consensus concernant la définition du terrorisme (Weinberg, Pedahzur, et Hirsch-Hoefler, 2004). Dans le contexte de cette étude, le passage à l'acte violent dans le but de contraindre l'État ou de terroriser la population et à des fins politiques, philosophiques, idéologiques, raciales, ethniques, religieuses ou autre sera considéré comme terrorisme. Cette définition sera utilisée comme critère d'exclusion pour les programmes antiterroristes visant à empêcher les attentats.

La radicalisation est un processus entamé, notamment par des individus ou des groupes. À l'échelle sociétale, nous utilisons plutôt le concept de *polarisation sociale* qui est la division progressive de la société et de l'environnement social en différents groupes et sous-groupes dont l'identité est basée sur l'exacerbation de caractéristiques opposées liées à des concepts essentiels comme le sexe, la race, la religion, les opinions politiques, etc. (CPN-PREV, 2020).

Par prévention, nous entendons tous les efforts visant à réduire ou à supprimer les conditions de risque qui peuvent rendre l'individu ou le groupe plus vulnérable à l'extrémisme violent ou qui le poussent vers la récidive quand il s'agit d'une personne qui a déjà recouru à la violence ou qui appartient à un groupe extrémiste. Sur la base de modèles de santé publique, la prévention peut cibler la population générale non identifiée comme étant à risque (prévention primaire), les individus ou les groupes considérés à risque ou qui se retrouvent dans les étapes initiales du processus de radicalisation menant à la violence (prévention secondaire) ou les individus ou les groupes qui sont déjà engagés dans ce processus dans les phases finales, qui appartiennent à un groupe extrémiste ou qui ont commis des actes associés à l'extrémisme violent (prévention tertiaire). Dans le cas de programmes de PEV, nous mettons également l'accent sur la distinction entre les programmes de prévention primaire et les programmes de prévention primaire ciblés. Ces derniers, bien qu'universels, ciblent une communauté spécifique.

Dans le contexte de cette étude, nous utilisons les concepts de *prévention de la radicalisation menant à la violence* et *prévention de l'extrémisme violent* comme des synonymes en privilégiant ce dernier et son acronyme (PEV). En revanche, nous distinguons les mesures de prévention de l'extrémisme violent des mesures de contreterrorisme. Si les mesures de prévention ont pour cible les individus vulnérables à l'extrémisme violent, les mesures de contreterrorisme, elles, visent les menaces à la sécurité et l'empêchement ou la dissuasion des attentats. Arce et Sandler (2005) font la différence entre les mesures de contreterrorisme proactives et défensives. Les premières sont souvent l'action directe de l'État ou ses agents contre les terroristes ou leurs commanditaires (destruction de camps d'entraînement terroristes, représailles contre un État sponsor, infiltration de groupes terroristes, etc.). En revanche,

les mesures défensives « visent à dissuader un attentat en rendant son succès plus difficile ou en augmentant les conséquences négatives probables pour l'auteur », comprenant l'érection de barrières technologiques, la fortification de cibles potentielles, la sécurisation des frontières, etc. (Arce et Sandler, 2005, p. 184).

Finalement, nous utilisons la définition d'évaluation utilisée par le groupe des Organisation des Nations unies pour l'évaluation : Une évaluation a pour objet d'apprécier, de manière aussi systématique et impartiale que possible, une activité, un projet, un programme, une stratégie, une politique, un sujet, un thème, un secteur, un domaine opérationnel ou une performance institutionnelle. Elle permet d'analyser le niveau de réussite, à la fois en termes de résultats escomptés et inattendus, en examinant la chaîne des résultats, les processus, les facteurs contextuels et le lien de causalité. Elle s'appuie pour cela sur des critères tels que la pertinence, l'efficacité, l'efficience, l'impact et la durabilité. Une évaluation doit fournir, à partir d'éléments démontrables, des renseignements crédibles et utiles, qui permettent d'intégrer en temps voulu les conclusions, recommandations et enseignements dans le processus décisionnel des organisations et des parties prenantes (UNEG, 2016, p. 10).

B2 CRITÈRES D'INCLUSION

Nous avons adopté des critères d'inclusion et d'exclusion qui maximisent l'inclusivité afin d'augmenter la probabilité de trouver des études malgré la variabilité quant aux cadres méthodologiques et théoriques. Les critères suivants sont un résumé des règles utilisées afin de juger un document scientifique comme admissible.

Nous avons ciblé toutes les études basées sur des données probantes primaires⁵⁴ qui ont évalué des programmes de PEV, publiées jusqu'en décembre 2019. Ces programmes ont eu pour but de réduire ou supprimer les conditions de risque qui peuvent rendre un individu ou un groupe plus vulnérable à l'extrémisme violent ou qui le poussent vers la récidive⁵⁵. Dans l'optique de la définition d'évaluation utilisée plus haut, nous avons inclus toutes les études qui ont eu pour but d'apprécier ou juger un programme, un projet ou une stratégie de PEV, y compris les études qui ne mentionnent pas directement le mot « évaluation ».

La population cible de ces études devait être conformée par des adultes. Nous ciblons, par conséquent, toutes les évaluations des programmes de PEV primaire, secondaire et tertiaire⁵⁶ qui essayent de modifier les attitudes, les émotions ou les comportements de ces individus ou de ces groupes, des familles ou de l'entourage de ces individus et des personnes intervenantes travaillant à ce sujet. En revanche, les évaluations de programmes qui prennent en charge les victimes directes ou indirectes⁵⁷, les évaluations des mesures de contreterrorisme ainsi que les études qui évaluent les stratégies continentales ou font un jugement global sur une approche continentale sont exclues.

Par conséquent, l'unité d'analyse est constituée par les études publiées et non par les publications. Une même publication peut faire état de plus d'une étude. On a considéré qu'il y avait plus d'une étude dans une publication lorsqu'il y avait plus d'un échantillon qui était analysé indépendamment et qui avait des résultats indépendants.

Mis à part les trois niveaux de prévention, il n'existe pas d'autre critère de classification de programmes à utiliser. Les variables à considérer ont été, par conséquent, décrites à partir de la comparaison de ces trois niveaux de prévention.

Les études devaient être rédigées en anglais, en français ou en espagnol (langues parlées par les membres de l'équipe de recherche).

Si toutes ces conditions étaient remplies, nous n'avons imposé aucune restriction additionnelle quant aux caractéristiques méthodologiques des études.

⁵⁴ Les données secondaires, pour leur part, désignent les données collectées par une personne autre que la personne chercheuse ou son équipe. Les exemples de sources de données secondaires dans les sciences sociales sont les recensements, les données collectées par les ministères, les registres organisationnels et d'autres données collectées qui avaient, à l'origine, d'autres objectifs de recherche.

⁵⁵ Voir les définitions clés utilisées dans la section précédente.

⁵⁶ Idem.

⁵⁷ Les familles des individus ayant entamé ce processus peuvent être considérées comme victimes indirectes des groupes extrémistes. Cependant, nous comprenons les « victimes » ici comme les personnes et leurs familles qui ont été la cible d'attentats, de tentatives d'attentat ou d'actions violentes de la part de groupes extrémistes.

B3 VARIABLES CODÉES

Chaque document scientifique a été codé en fonction d'une grille de codage global et d'un outil d'évaluation de la qualité méthodologique.

a) Grille de codage global

La grille de codage suivante a été développée pour coder puis agréger les données issues des documents scientifiques. Le codage a été effectué par une équipe d'auxiliaires de recherche sur la base de la liste suivante :

Dimension	
Variable	Définition opérationnelle
Description générale	
Auteur/auteure	—
Pays	Pays du programme de prévention
Révision par des pairs	Si les documents scientifiques ont été évalués par d'autres chercheuses et chercheurs (aveuglement). C'est le cas, notamment, des revues scientifiques.
Financement de la recherche	Si les chercheuses et chercheurs mentionnent les sources de financement (Oui, préciser la source de financement).
Conflit d'intérêts	Si les chercheuses et chercheurs mentionnent leurs conflits d'intérêts.
	Liste de conflits d'intérêts mentionnés.
	Liste de conflits d'intérêts non mentionnés.
Identification des auteures et auteurs des études retenues	
Genre	Genre de l'auteur ou auteure
Pays de provenance	Pays de provenance de l'auteur ou auteure
Discipline	Discipline de l'auteur ou auteure
Profession	Profession de l'auteur ou auteure
Nombre de publications comme seul auteur	Nombre de publications comme seul auteur/auteure dans le domaine de la sécurité
Nombre de publications comme co-auteur	Nombre de publications comme co - auteur/auteure dans le domaine de la sécurité
Nombre de publications dans la base de données	Nombre de publications dans la base de données de cette revue systématique
Région de la première publication	Région géographique de la première publication de l'auteur ou auteure
Niveau de prévention	
Primaire	Tous les efforts visant à réduire ou à supprimer les facteurs de risque ou à encourager les facteurs de protection et qui ciblent la population générale non identifiée comme étant à risque. C'est un type de prévention universelle. Les campagnes de sensibilisation, par exemple.
Primaire ciblée	Tous les efforts visant à réduire ou à supprimer les facteurs de risque ou à encourager les facteurs de protection et qui ciblent une communauté spécifique et non identifiée comme étant à risque. Exemple : les programmes de prévention universelle auprès de communautés musulmanes (Channel).
Secondaire	Tous les efforts visant à réduire ou à supprimer les facteurs de risque ou à encourager les facteurs de protection et qui ciblent les individus ou les groupes considérés à risque ou qui se retrouvent dans les étapes initiales de ce processus.
Tertiaire	Tous les efforts visant à réduire les facteurs qui encouragent la récidive dans le cas des individus ou des groupes qui se retrouvent dans les phases finales de ce processus, qui appartiennent à un groupe extrémiste ou qui ont commis des actes associés à l'extrémisme violent ou au terrorisme. Ce type de prévention vise également la réinsertion en société de ces individus ou de ces groupes.
Générale	Ce n'est pas clairement indiqué dans l'étude.

Type d'extrémisme	
Extrême gauche (ou synonymes)	Le programme ou projet cible directement ce type d'extrémisme (c'est clairement indiqué dans le document).
Extrême droite (ou synonymes)	Le programme ou projet cible directement ce type d'extrémisme (c'est clairement indiqué dans le document).
Djihadisme (ou synonymes)	Le programme ou projet cible directement ce type d'extrémisme (c'est clairement indiqué dans le document).
Anarchisme (ou synonymes)	Le programme ou projet cible directement ce type d'extrémisme (c'est clairement indiqué dans le document).
Autre	Le programme ou projet cible n'importe quel autre type d'extrémisme qui ne s'ajuste pas aux autres descriptions.
Non spécifique	Le programme ou projet cible tout type d'extrémisme (c'est clairement indiqué dans le document). C'est le cas, notamment, de programmes de prévention primaire.

Type de modèle d'évaluation	
Impact (ou sommative)	L'évaluation d'impact répond à la question « qu'est-ce qui a fonctionné ? », c'est-à-dire qu'elle s'interroge sur les effets que l'intervention a eus sur les participantes et participants et si ces effets correspondent aux objectifs proposés. Ce sont les évaluations qui évaluent la façon dont une intervention contribue à atteindre un résultat ou un objectif. La contribution peut être intentionnelle ou non, positive ou négative, à long ou à court terme. Les évaluations d'impact tentent d'identifier un lien clair entre les causes et les effets et d'expliquer comment l'intervention a fonctionné et pour qui.
Processus (ou formative)	L'évaluation de processus répond aux questions « Pourquoi est-ce qu'elle fonctionne ? » « Comment est-ce qu'elle fonctionne ? » et « Comment est-ce que nous pouvons améliorer ce processus ? ». Une évaluation de processus se centre ainsi sur les facteurs qui déterminent ou conditionnent la mise en place des activités d'intervention et sert à comprendre les évolutions, les changements qui ont lieu au cours d'un projet. Cette évaluation de processus peut avoir lieu peu après le démarrage de l'intervention (évaluation formative), durant l'intervention (évaluation du processus) ou au milieu de l'intervention (évaluation à mi-parcours).
Évaluation d'extrants (ou de résultats)	Étude conduite en fin d'action (ou à la fin d'une étape de cette action) pour déterminer dans quelle mesure les réalisations escomptées ont été atteintes (<i>output evaluation</i>).
Audit (contrôle)	Activité de contrôle en matière de qualité, exercée de façon objective et indépendante, et destinée à améliorer les opérations d'une organisation et à en accroître la valeur. L'audit aide une organisation à atteindre ses objectifs grâce à une approche systématique et rigoureuse pour constater et améliorer l'efficacité de la gestion des risques, du contrôle et des processus de gouvernance.
Suivi (monitoring)	Processus continu de collecte systématique d'informations, selon des indicateurs choisis, pour fournir aux gestionnaires et aux parties prenantes d'une action de développement en cours, des éléments sur les progrès réalisés, les objectifs atteints et l'utilisation des fonds alloués.
Autre type d'évaluation	Tout autre type d'évaluation.

Type d'évaluateur et d'évaluatrice	
Évaluation interne	Évaluation réalisée par celles et ceux qui ont la responsabilité de concevoir et de mettre en œuvre une action ou programme. Évaluation conduite par un service ou des personnes qui dépendent des responsables de l'action ou programme chez le bailleur de fonds, chez ses partenaires ou au sein des organisations chargées de la mise en œuvre.
Évaluation conjointe (<i>partenariale</i>)	Évaluation à laquelle participent différents organismes bailleurs ou leurs partenaires, sans compter les bénéficiaires ou les personnes intervenantes.
Évaluation participative	Méthode d'évaluation selon laquelle toutes les parties prenantes (y compris les bénéficiaires, les personnes intervenantes et chercheuses) collaborent pour concevoir et conduire une évaluation et en tirer les conclusions.
Évaluation externe (évaluation indépendante)	Évaluation d'une action ou un programme conduit par des services ou des personnes extérieures au bailleur de fonds et à l'organisation responsable de la mise en œuvre. Évaluation d'une action de développement conduite par des services ou des personnes non liés aux responsables de la conception et de la mise en œuvre de l'action de développement.

Design ou approche méthodologique : En fonction de l'approche globale	
Quantitative	Études qui utilisent des variables quantifiables et dont le système de cueillette de données direct (observations) ou indirect (sondage) est quantitatif et l'analyse est statistique : des sondages, des observations codées en chiffres, etc.
Qualitative	Études qui utilisent des méthodes de cueillette de données et d'analyse qualitatives : observations participantes, ethnographies, entretiens, groupes de discussion, etc.
Mixte	Études qui utilisent les deux types de méthode (quali-quant)
Autres	Toute autre approche globale.

Design ou approche méthodologique : En fonction de la manipulation de variables	
Expérimentale	La stratégie expérimentale implique la manipulation active de la variable indépendante, c'est-à-dire que la personne chercheuse choisit arbitrairement les valeurs de la variable indépendante (l'intervention, par exemple) et les applique à différents groupes de sujets afin de tester une relation de causalité. Les mesures sont prises en au moins deux temps (avant et après l'intervention) et dans plus d'un groupe. Ce type d'expérience se fait normalement avec un groupe témoin. Les personnes participantes de ce groupe comme celles du groupe d'expérimentation ont été sélectionnées de façon aléatoire.
Quasi expérimental	Ce modèle a l'intention de tester la causalité de l'intervention par rapport à ces effets avec une prise de mesures avant et après l'intervention, mais contrairement à la stratégie expérimentale, soit qu'il n'y a pas de groupe témoin, soit que les groupes ne sont pas formés au hasard, mais utilisent des groupes naturels, intacts ou déjà formés.
Autre	

Design ou approche méthodologique : En fonction des personnes participantes	
Groupe témoin	Un groupe de sujets ressemblant étroitement au groupe de traitement pour de nombreuses variables démographiques, mais ne recevant pas l'intervention et servant ainsi de groupe de comparaison lors de l'évaluation des résultats du traitement.

Design ou approche méthodologique : Mesures répétées	
Mesures répétées	Les designs à mesures répétées dans les évaluations de ces programmes sont basés sur plusieurs observations des mêmes sujets à au moins deux moments différents.
Post-évaluation	Évaluation d'un programme à un seul moment, une fois que celui-ci est terminé ou une fois qu'un cycle d'intervention est terminé.

Design ou approche méthodologique : En fonction du nombre de variables indépendantes	
Simple	Une seule variable indépendante.
Complexe ou factoriel	Plus d'une variable indépendante.

Design ou approche méthodologique : En fonction du nombre de variables dépendantes	
Simple	Une seule variable dépendante.
Complexe ou factoriel	Plus d'une variable dépendante.

Outils de collecte de données	
Sondages	C'est une technique de collecte de données quantitative qui se fait par l'intermédiaire d'un ensemble de questions standardisées adressées à un échantillon pour connaître des états d'opinion ou diverses questions de fait.
Entretiens	C'est une technique de collecte de données qualitatives utilisée dans les sciences sociales, qui nous permet de connaître et d'étudier les opinions et les attitudes d'une personne par rapport à un sujet spécifique et par le biais d'un modèle de conversation.
Groupes de discussion	C'est une technique de collecte de données qualitatives utilisée dans les sciences sociales, qui nous permet de connaître et d'étudier les opinions et les attitudes d'un groupe de personnes par rapport à un sujet spécifique.
Observation	L'observation est une technique de collecte de données dans le cadre d'une recherche qualitative ou quantitative. Dans le cas de la recherche qualitative, son but est de se familiariser avec un groupe particulier d'individus (comme un groupe religieux, professionnel, sous-culturel ou une communauté particulière) et leurs pratiques par un engagement intensif avec les personnes dans leur environnement culturel, généralement sur une longue période. Dans le cas de la recherche quantitative, il s'agit de l'utilisation d'une grille d'observation préétablie et dont les données collectées seront quantifiées et analysées statistiquement.
Autres	

Type d'action ciblée	
Évaluation du plan ou de la stratégie nationale	Évaluation de l'ensemble des actions mises en place à travers une stratégie nationale.
Évaluation partielle du plan ou de la stratégie nationale	Évaluation d'une partie des actions mises en place à travers une stratégie nationale dans un échantillon, un secteur ou une région spécifique.
Évaluation de projet	Évaluation d'une action de prévention individuelle conçue pour atteindre des objectifs spécifiques avec des ressources et un plan de travail déterminés.

Échantillon	
Participantes et participants à un groupe d'intervention	Nombre de personnes participant à un groupe d'intervention.
Participantes et participants à un groupe témoin	Nombre de personnes participant à un groupe témoin.

Population cible	
Individus directement impliqués	S'applique lorsqu'il y a des interventions auprès d'individus spécifiques et, notamment, dans le cas de la prévention secondaire ou tertiaire, c.-à-d. quand ils ont déjà entamé le processus ou quand ils se trouvent à la fin de la trajectoire.
Familles	Lorsqu'un service est offert aux familles de personnes associées à l'extrémisme violent.
Communautaire	Lorsque l'intervention implique le fait de travailler avec l'entourage large (excluant la famille) de l'individu à l'échelle locale (dans le cas de la prévention primaire, c'est la majorité des cas).
Groupe sociétal	Lorsque l'intervention implique le fait de travailler avec un groupe sociétal spécifique, mais pas avec la société dans sa totalité (jeunes, musulmans, femmes, etc.).
Société	La société est la cible de l'intervention (prévention primaire ou universelle). C'est le cas, notamment, de campagnes de sensibilisation.
Intervenantes et intervenants	Lorsque l'intervention a également pour but de travailler avec n'importe quelle personne qui a un contact direct avec les participantes et participants de l'intervention.
Gouvernement	Quand l'intervention implique le renforcement de capacités au sein d'un organisme gouvernemental afin d'améliorer ses capacités de prévention.
Milieu cible du programme de prévention	
Milieu communautaire	Lorsque l'intervention implique le fait de travailler avec l'entourage large (excluant la famille) de l'individu à l'échelle locale.
Milieu de la sécurité	Lorsque l'intervention cible la population des forces de l'ordre et des forces armées.
Milieu éducationnel primaire et secondaire	Lorsque l'intervention cible la population du secteur éducationnel primaire et secondaire.
Milieu éducationnel universitaire	Lorsque l'intervention cible la population du secteur éducationnel post secondaire.
Milieu de la justice	Tous les organismes associés au système de justice (justice de mineur et tribunaux par exemple) à l'exception du système correctionnel et le secteur de la sécurité
Gouvernement	Toutes les institutions du gouvernement à l'exception du milieu éducationnel, de la santé et correctionnel
Milieu culturel	
Milieu correctionnel	Lorsque l'intervention cible la population du secteur correctionnel fermé, semi-ouvert et le système de probation
Secteur privé	Lorsque l'intervention cible la population des organismes à but lucratif.
Milieu de la santé	Toutes les institutions du milieu de la santé physique ou mentale.
Autre	
Type d'indicateurs utilisés ou résultats obtenus	
Facteur ou variable, de nature quantitative ou qualitative, qui constitue un moyen simple et fiable de mesurer et d'informer des changements liés à l'intervention.	
Directs	Des indicateurs qui mesurent directement la radicalisation, l'extrémisme violent ou les sympathies pour ces phénomènes.
Indirects	Des indicateurs non directement liés à l'extrémisme violent ou aux sympathies pour l'extrémisme violent. Par exemple, estime de soi, leadership, etc.
Indicateurs utilisés ou résultats obtenus	Liste d'indicateurs rapportés.
Types d'effets	
Liste d'effets positifs et négatifs rapportés	Positifs
	Négatifs
	Autres
Limites de la recherche	
Limites	Les auteures et auteurs mentionnent-ils les limites de la recherche ?
Types de limites indiqués	Liste des limites rapportées.

B4 Outil d'évaluation de la qualité méthodologique

En addition à ce groupe de variables, nous avons utilisé le *Mixed Methods Appraisal tool (MMAT)* (Hong et al., 2018 ; Hong et Pluye, 2019) comme outil d'évaluation de la qualité méthodologique des documents scientifiques retenus dans cette revue systématique. À la différence d'autres outils d'évaluation, le MMAT permet d'évaluer, au sein du même instrument, des études qualitatives, quantitatives descriptives, expérimentales, quasi expérimentales et mixtes qui étaient les variétés d'études que nous espérions retrouver lors de la recherche. En raison de notre intérêt à répertorier toutes les méthodologies d'évaluation de programmes de PEV, nous n'avons pas utilisé le MMAT comme un critère d'inclusion

de documents scientifiques. Il a servi, au contraire, seulement pour déterminer la qualité des méthodologies utilisées dans le domaine.

Le MMAT se compose de 25 variables réparties en 5 groupes représentant les cinq types d'études mentionnés ci-dessus. Chaque étude peut obtenir un score entre 0 et 5. Cependant, dans le cas des études qui utilisent des méthodologies mixtes, il faut aussi coder les variables associées à chaque type d'étude concerné. Une étude qui utilise des méthodologies mixtes peut potentiellement obtenir un score entre 0 et 25.

B5 Stratégie de recherche de la littérature

Nous avons utilisé la liste de mots clés suivants en anglais et en français :

ANGLAIS	(Extremi* OR Radicali* OR "Violent Extrem*" OR Indoctrinat* OR Terrori* OR "Homegrown Terror*" OR "Homegrown Threat*" OR "Radical Islam*" OR "Islamic Extrem*" OR "Religious Extrem*" OR Fundamentalis* OR Jihad* OR Islam* OR Salaf* OR "Lone wol*" OR "lone-wol*" OR "lone actor*" OR "foreign fight*" OR Returne* OR "White Supremacis*" OR "Neo-Nazi" OR "Right Wing" OR "Right-wing Extrem*" OR "far right" OR Fascis* OR "Left-wing Extrem*" OR "Left Wing" OR Anti-Semitis* OR Antifa* OR Anarch* OR "Eco-terror*" OR "Al Qaida-inspired" OR "ISIS-inspired" OR "Anti-Capitalis*" OR Incel* OR "Al Qaeda" OR ISIS OR ISIL) AND (Prevent* OR interven* OR respon* OR policy OR policies OR program* OR strategy* OR initiative* OR assess* OR eval* OR procedur* OR effect* OR *success* OR reduc* OR treat* OR counterterror* OR "counter-terror*" OR "de-radicali*" OR deradical* OR disengag* OR detect* OR "countering violent extrem*" OR CVE OR PVE OR Reint* OR Rehabilitat*) NOT (Cancer OR Disease OR hematoma OR "heart disease" OR "heart failure" OR cardiovascular OR "vortex generator*" OR "heat transfer" OR "bone" OR "fracture healing" OR "bone density" OR epilepsy OR "multiple sclerosis" OR Femin*)
FRANÇAIS	(Extremi* OR Radicali* OR "Extrem* Violent" OR Endoctrin* OR Terrori* OR " Terror* Domestique" OR "Islam* Radical" OR "Extrem* Islam*" OR "Extrem* Relig*" OR Fundamentalis* OR djihad * OR Islami* OR Salaf* * OR "Loup* solitaire*" OR "acteur solitaire *" OR (combattant* AND (étranger* OR terroriste*)) OR "Extrême droite" OR Suprémac* OR "Néo-Nazi" OR Néonazi* OR Fachis* OR "Extrem* Gauch" OR Antifa* OR Anti-Semitis* OR Anarch* OR "Eco-terror*" OR Incel* OR "Al Qaeda" OR ISIS OR ISIL) AND (Prevent* OR interven* OR repon* OR politique* OR program* OR stratégie* OR initiative* OR eval* OR procedur* OR effet* OR effect* OR succès OR réussi* OR résultat* OR reduc* OR traitem* OR contreterror* OR "contre-terror*" OR "de-radicali*" OR deradical* OR disengage* OR CVE OR PVE OR Reintegr* OR Rehabilitat* OR reinsert*) NOT (Cancer OR Maladi* OR Hématom* OR "Maladi* cardia*" OR "Insuffisan* cardia*" OR Cardiovasculaire* OR "Générat* de tourbillon*" OR "Transfer* de chaleur*" OR Os OR "Consolid* de fractur*" OR "Densit* osseu*" OR Épileps* OR "Scléro*" OR femin*)

À partir de critères d'inclusion et d'exclusion et de ces mots clés, nous avons utilisé quatre stratégies de recherche :

- Revue de la littérature scientifique.
- Revue de la littérature grise.
- Révision d'autres recensions ou revues de la littérature fréquemment citées et boule de neige.
- Communications avec des auteures et auteurs.

a) Littérature scientifique

Nous avons fait appel à un bibliothécaire expert dans le domaine des sciences sociales et humaines qui, sur la base de nos critères, a réalisé une recherche sur les bases de données suivantes :

ABI/Inform Global
Academic Search Complete
ATLA Religion Database
Canadian Business et Current Affairs Complete
Communication Abstracts
Canadian Public Policy Collection
Canadian Research Index
Education Source
ERIC
Erudit / Persee
FRANCIS

International Political Science Abstracts
Medline
[OpenGrey.eu](https://opengrey.eu)
PAIS Index
Political Science Complete
ProQuest Dissertations et Theses Global
PsycINFO
Sociological Abstracts
Sociological Index
Web of Knowledge.

Ces bases de données nous ont permis de repérer non seulement des articles scientifiques ou thèses, mais aussi un nombre important de littérature grise et de communications lors de congrès. Nous avons aussi obtenu un accès à la base de données d'une revue systématique du Réseau des Praticiens Canadiens pour la Prévention de la Radicalisation et de l'Extrémisme Violent (RPC-PREV) (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021). Les deux bases de données ont été fusionnées.

b) Littérature grise

Afin de réduire le « biais de publication » (Rothstein et al., 2005), nous avons effectué une recherche approfondie de la littérature grise en utilisant Google. En outre, les sites Web des organisations travaillant dans le domaine de la radicalisation et de la lutte contre l'extrémisme violent ont été recherchés manuellement pour trouver d'autres documents. Ces organismes ont été sélectionnés à partir d'une cartographie de pôles d'expertises en PEV réalisée par la Chaire UNESCO-PREV de façon préalable (<https://chaireunesco-prev.ca/lereseau/cartographie/>). D'autres organismes ont été inclus au cours de la recherche. Le Tableau 32 dresse la liste complète des organismes sélectionnés.

Tableau 32. Liste des organismes sélectionnés

Academy on Human Rights and Humanitarian Law at American University's Washington College of Law	Center for Strategic and International Studies (CSIS)
Afghanistan Justice Organization	Center on Global Counterterrorism cooperation (CGCC)
AfPak programme Afghanistan/Pakistan (PSF)	Center on International Cooperation at New York University
Againstviolentextremism.org	Centre africain d'Etudes Internationales, Diplomatiques, Economiques et Stratégiques, en abrégé (CEIDES)
Akin Gump Strauss Hauer & Feld LLP	Centre for Environment, Human Rights and Development (CEHRD - Nigéria)
Alliance for Peacebuilding – Monitoring and Evaluation of CVE	Centre for Environmental Education and Development (CEED - Nigéria)
Alternative espaces citoyens (AEC - Niger)	Centre for Peace And Advancement (CEPAN - Nigéria)
Alternative to Violence Project	Centre for Research and Evidence on Security Threat (CREST)
American Bar Association Rule of Law Initiative (ABA ROLI)	Centre for the Advocacy of Justice and Rights (CAJR)
Amicale université populaire (Tchad)	Centre pour la Gouvernance Democratique
Amicus Legal Consultants	Century Foundation
AML Solutions International	Charity & Security Network
Amnesty International	Children and Young People Living for Peace (Nigéria)
Anti-Defamation League	Christian Foundation for Social Justice and Equity (CFSJE - Nigéria)
APO.org	Civipol
Asser Institute	Cleen Foundation (Nigéria)
Association burkinabé d'action communautaire (ABAC-ONG - Burkina Faso)	Clingendael – Netherlands Institute of International Relations
Association des jeunes juristes et sympathisants de Sikasso (AJJSS - Mali)	Club UNESCO de l'Université Abdou Moumouni (CUAM - Nigéria)
Association for Progressive Communications	CODE PAKISTAN
Association jeunesse pour la paix et la non-violence (AJPNV - Tchad)	Collectif des organisations de défense des droits de l'homme et de la démocratie (CODDHD - Niger)
Association of Francophone Supreme Courts (AHJUCAF)	Comité Interministériel de prévention de la délinquance et de la radicalisation (CIPDR)
Association pour l'enseignement coranique et la protection des enfants mouhadjirine (AECPEM - Tchad)	Commission Européenne
Association pour le dialogue entre les jeunes de diverses religions (ADJR - Tchad)	Community Motivation and Development Organization (CMDO)
Association rayons de soleil (Cameroun)	Community Policing Partners for Justice, Security & Democratic Reform (Nigéria)
Association tchadienne pour la promotion et la défense des droits de l'homme (ATPDH - Tchad)	Conflict Resolution Trainers Network (CROTINN - Nigéria)
Attah Sisters Helping Hand Foundation (ASHH - Nigéria)	Council of Europe
Baker & McKenzie	COWI
Bangladesh Enterprise Institute (BEI)	Danish Ministry of Defence (Broad Peace and Stabilisation Fund)
Bangladesh Institute of Peace and Security Studies (BIPSS)	Danish security and intelligence service
Better World Campaign	Defence, Australian Government
Bipartisan Policy Center	Design Monitoring and Evaluation for Peacebuilding
Blumont.org	Development Initiative of West Africa (DIWA - Nigéria)
Brennan Center for justice	Development, Education and Advocacy Resources for Africa (DEAR Africa - Nigéria)
Brookings Institution	Djamah-Afrik (Tchad)
Burkina Faso CRADHE	Dorwood Consultancy
Cadre africain de coopération civilo-militaire (CCCM- Niger)	East Africa Judges' and Magistrates' Association (EAJMA)
Care Fronting (Nigéria)	
Center for Evidence Based Crime Policy CEBCP	
Center for prevention of radicalization leading to violence	

Economic Community of West African States (ECOWAS)
 Educateagainsthate.com
 Education and Community Development
 EducommunicAfrik (Burkina Faso)
 Emergency Preparedness and Response Team (JDPC- Nigéria)
 Equal Access International
 EU Agency for Fundamental Rights
 European Counter-Radicalization and de-radicalization
 European Judges Training Network (EJTN)
 Exit Sweden
 Fantsuam Foundation (Nigéria)
 Federation burkinabé des associations, centres et clubs UNESCO (FBACU- Burkina Faso)
 Fondation Hirondelle (Niger et Mali)
 Ford Foundation
 Fourth Freedom Forum
 French Ministry of Interior Publications Database
 Friedrich Naumann Foundation (South Asia)
 Geneva Centre for Security Policy (GCSP)
 Georgetown University Center for Security Studies
 German National Center for Crime Prevention
 Global Center on Cooperative Security GCCS
 Global Community Engagement and Resilience Fund (GCERF)
 Global Counter Terrorism Forum (GCTF)
 Global Counter Terrorism Forum Violent Extremism (Hedayah)
 Global Initiative against Transnational Organized Crime
 Global Partnership for the Prevention of Armed Conflict
 Graduate Institute of International and Development Studies
 GW Program on Extremism
 Henry L. Stimson Center
 Hope for the Needy Association (HOFNA - Cameroun)
 Horn of Africa (HoA) programme (PSF)
 Human Rights First
 Human Rights Institute at Columbia University Law School
 Human security collective
 ICF
 IDP Goods (Cameroun)
 Impact Europe
 Inganta Rayuwa Peace Network (Nigéria)
 Insan Foundation
 Institut national de la statistique et des études économiques (INSEE)
 Institut of Security Studies
 Institute for Inclusive Security

Institute for Justice and Reconciliation
 Institute for Social Policy and Understanding
 Institute for strategic dialogue (ISD)
 Integrity research and consultancy
 Integrityglobal.com
 Interfaith Council of Muslim and Christian Women's Associations (Nigéria)
 Intergovernmental Authority on Development (IGAD)
 International Association of Chiefs of Police (IACP)
 International Centre for Counter-Terrorism – The Hague (ICCT)
 International Centre for Peace, Charities and Human Development (INTERCEP - Nigéria)
 International Centre for the study of Radicalisation (ICSR)
 International Centre of Excellence for Countering Violent Extremism
 International Crisis Group
 International Institute for Justice and the Rule of Law (IIJ)
 International Monetary Fund (IMF)
 International Organization for Judicial Training (IOJT)
 International Peace Institute (IPI)
 International Republican Institute (IRI -Niger, Mali)
 Interpol
 Islamabad Policy Research Institute
 Islamic Counselling Initiatives of Nigeria (ICIN - Nigéria)
 Istituto Affari Internazionali
 Kecosce
 Kingsfaith Development and Youth Empowerment Initiative (Nigéria)
 Knowledge Platform Security& Rule of Law
 Leadership Initiative for Transformation and Empowerment (LITE- Africa - Nigéria)
 Leiden university
 Media Women for Peace (Cameroun)
 Ministry of Foreign Affairs of Denmark
 Moonshot
 Mouvement des jeunes pour le développement et l'éducation citoyenne (MOJEDEC - Niger)
 Nahdatul Ulama (NU)
 National Consortium for the Study of Terrorism and Responses to Terrorism (START)
 National Counterterrorism Center
 National Endowment for Democracy
 NATO Science for Peace and Security Program
 Neem Foundation (Nigéria)
 New Era Educational and Charitable Support Foundation (Nigéria)
 North East Youth Initiative for Development (Nigéria)
 Norwegian Ministry of Foreign Affairs

Observer Research Foundation (ORF)	U.S. Agency for International Development (USAID)
Office of the United Nations High Commissioner on Human Rights (OHCHR)	UiO C-REX - Center for Research on Extremism
ONG Adkoul (Niger)	UK College of Policing
ONG Jeunesse-enfance-migration-développement (JMED - Niger)	UK Home Office Research Database
Open Society Foundation	UK Ministry of Defence
Organisation for Security and Cooperation in Europe (OSCE)	UN Counter-Terrorism Implementation Task Force (CTITF)
Organisation pour la réflexion, la formation et l'éducation à la démocratie et au développement (ORFED - Mali)	UN Office of the High Commissioner for Human Rights
OXFAM	UN Office of the Special Adviser on Africa
PAIMAN Alumni Trust	UN Office on Drugs and Crime (UNODC)
Pak Institute for Peace Studies Pvt Ltd. (PIPS)	UN Security Council Counter-Terrorism Committee Executive Directorate (CTED)
Peace and Stabilisation Fund (Danemark)	UN Women
Peace Empowerment Foundation (Nigéria)	UNESCO
Peace Initiative Network (PIN) (Nigéria)	Union Européenne
Prevention of and Fight against crime programme of the European union European commission	United Nations
RAND Corporation	United Nations Association – UK
Regional Center for Strategic Studies	United Nations Development Programme (UNDP)
Réseau de Réflexion Stratégique sur la Sécurité au Sahel	United Nations Foundation
Réseau panafricain pour la paix, la démocratie et le développement (REPPADD)	United Nations Office for Drugs and Crime's Terrorism Prevention Branch (UNODC)
Royal Canadian Mounted Police (RCMP)	United Nations Peacebuilding Support Office (PBSO), Niger et projet régional
Royal United Services Institute (RUSI)	United States Institute of Peace (USIP)
SaferWorld	University Of Cambridge (institute of criminology)
Salesforce	US Department of Homeland Security
Search for common Ground	US National Criminal Justice Reference Service
South Asian Association for Regional Cooperation (SAARC)	Violence Prevention Network (Germany)
Stockholm International Peace Research Institute (SIPRI)	West Africa Network for Peacebuilding (WANEP)
Stop-djihadisme (France)	Women Against Violent Extremism (WAVE - Nigéria)
Stoppingviolentextremism.org	Women and Girl Child Rescue and Development Initiative (Nigéria)
Strong Cities Network (SCN)	Women in International Security (WIIS)
Tabara Youth Transformation Initiative (TYTI- Nigéria)	World Affairs Council
Taimako Community Development Initiative (Nigéria)	World Bank
Tech Against Terror	World Organization for Resource Development and Education (WORDE)
The Campbell Collaboration	Youth Initiative Against Violence and Human Rights Abuse (YIAVHA - Nigéria)
The Global Observatory	Youth Justice Board
The John Sloan Dickey Center of International Understanding – Dartmouth University	Youth Progressive Association in Taraba (TYPA - Nigéria)
The Prevention Project	Youths for Peace Building and Development in Africa (YOUPEDA - Nigéria)
The Unity Initiative (TUI)	

c) Autres recensions ou revues de la littérature fréquemment citées

En plus des documents identifiés à l'aide de la stratégie de recherche décrite ci-dessus, nous avons comparé nos résultats avec des revues ou recensions fréquemment citées dans la littérature (Tableau 33).

Tableau 33. Revues systématiques et recensions de la littérature sur l'évaluation de programmes de prévention de la radicalisation et de l'extrémisme violents

Revue systématique ou recension de la littérature	Retenues par cette étude	Études exclues			
		CT*	SDP*	NE*	F*
Bellasio et al., 2018	28/48	7	3	2	8
Carthy et al., 2020	0/14	14			
Feddes et Gallucci, 2015	11/55	6	19	2	17
Gielen, 2017	25/73	4	38	3	3
Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021; Hassan, Brouillette-Alarie, Ousman, Savard et al., 2021	44/48	2	1		1
Madriaza et al., 2017; Madriaza et Ponsot, 2015	12/23	6	3		2
Mastroe et Szmania, 2016	16/43	7	14	1	5
Pistone et al., 2019	17/38	5	12	2	2
Pratchett et al., 2010	1/18	4	6		7
Taylor et Soni, 2017	1/7	5		1	

* : N = Nombre d'études qui mentionnent le nombre de participants ;

Total = Nombre total d'études identifiées

CT : Étude classifiée comme mesure antiterroriste, non directement liée à la prévention ou qui ne porte sur aucun programme spécifique.

SDP : Étude sans données primaires ou avec des données anecdotiques.

NE : Étude non évaluative.

F : Publications inaccessibles ou fusionnées avec une autre publication qui a utilisé le même échantillon et analyse.

Chaque document dans les langues sélectionnées que nous n'avions pas identifié a été ajouté à notre base de données. En addition à ces recensions, une stratégie de boule de neige a été réalisée à partir de la bibliographie des études retenues.

d) Communications avec les auteures et auteurs

Nous avons également consulté par courrier électronique 14 expertes et experts pour savoir s'ils connaissaient d'autres études.

B6 PROCÉDURE

Avant le début des travaux, il y a eu une période de formation pour les cinq auxiliaires pour clarifier les concepts et la méthodologie de travail. Nous avons utilisé deux bases de données bibliographiques. L'une provient d'une revue systématique semblable déjà faite par l'équipe RPC-PREV (Hassan, Brouillette-Alarie, Ousman, Kilinc et al., 2021) qui avait certains des mots clés en commun. Cette base de données allait jusqu'en janvier 2018. Une nouvelle revue bibliographique a été conduite en utilisant les critères déjà mentionnés. Pendant que le bibliothécaire faisait cette cueillette de documents scientifiques, les auxiliaires de recherche ont révisé la littérature grise disponible sur les sites web des organismes mentionnés ci-dessus. Dès que la cueillette de données de la littérature grise a été achevée, les bases de données ont été fusionnées et les doublons ont été éliminés. Les 14 personnes expertes ont été contactées pendant cette période.

Pour éliminer les études non admissibles, les auxiliaires de recherche et le chercheur principal ont passé au crible les titres et les résumés des documents identifiés lors de la recherche documentaire. Pendant cette première phase, l'ensemble de l'équipe a codé les 700 premiers documents afin d'assurer une certaine cohérence. Les désaccords ont été analysés et résolus. Cette phase a aussi servi de formation à l'équipe. Par la suite, deux personnes codeuses ont révisé chaque document. Pour s'assurer que l'accord entre les personnes évaluatrices était adéquat, un Kappa de Cohen a été calculé. Pendant ce codage initial, nous avons travaillé de façon itérative, c'est-à-dire que chaque couple a travaillé sur un nombre limité d'articles, ensuite le Kappa de Cohen a été calculé et, si le résultat était de moins de 0,6 (le minimum admissible), le couple révisait leurs désaccords. Si le résultat du Kappa était en dessus de 0,6, le couple continuait à coder la tranche suivante de documents. Le Kappa final a été de 0,86.

Le nombre total de publications sélectionnées était de 211, mais certaines publications faisaient état de plus d'une étude. En conséquence, le nombre total d'études retenues a été de 219. On a considéré qu'il y avait plus d'une étude dans une publication lorsqu'il y avait plus d'un échantillon qui était analysé indépendamment.

Nous avons utilisé le modèle PRISMA (<http://www.prisma-statement.org>) pour consigner les résultats des recherches documentaires dans un organigramme (voir la [page 27](#)).



Organisation
des Nations Unies
pour l'éducation,
la science et la culture



• « **Chaire UNESCO** en prévention de
• la radicalisation et de l'extrémisme violents »
• Université de Sherbrooke, Université Concordia
• et Université du Québec à Montréal